

UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie électrique et de génie informatique

CONCEPTION D'UN CIRCUIT
D'ÉTOUFFEMENT POUR
PHOTODIODES À AVALANCHE EN
MODE GEIGER POUR INTÉGRATION
HÉTÉROGÈNE 3D

Mémoire de maîtrise
Spécialité : génie électrique

Alexandre BOISVERT

Jury : Jean-François PRATTE (directeur)
Réjean FONTAINE (codirecteur)
Serge CHARLEBOIS

Sherbrooke (Québec) Canada

Janvier 2014



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-499-00307-2

Our file Notre référence

ISBN: 978-0-499-00307-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

RÉSUMÉ

Le Groupe de Recherche en Appareillage Médical de Sherbrooke (GRAMS) travaille actuellement sur un programme de recherche portant sur des photodiodes à avalanche monophotoniques (PAMP) opérées en mode Geiger en vue d'une application à la tomographie d'émission par positrons (TEP). Pour opérer dans ce mode, la PAMP, ou SPAD selon l'acronyme anglais (*Single Photon Avalanche Diode*), requiert un circuit d'étouffement (CE) pour, d'une part, arrêter l'avalanche pouvant causer sa destruction et, d'autre part, la réinitialiser en mode d'attente d'un nouveau photon. Le rôle de ce CE comprend également une électronique de communication vers les étages de traitement avancé de signaux. La performance temporelle optimale du CE est réalisée lorsqu'il est juxtaposé à la PAMP. Cependant, cela entraîne une réduction de la surface photosensible ; un élément crucial en imagerie. L'intégration 3D, à base d'interconnexions verticales, offre une solution élégante et performante à cette problématique par l'empilement de circuits intégrés possédant différentes fonctions (PAMP, CE et traitement avancé de signaux). Dans l'approche proposée, des circuits d'étouffement de $50\text{ }\mu\text{m} \times 50\text{ }\mu\text{m}$ réalisés sur une technologie CMOS 130 nm 3D Tezzaron, contenant chacun 112 transistors, sont matricés afin de correspondre à une matrice de PAMP localisée sur une couche électronique supérieure. Chaque circuit d'étouffement possède une gigue temporelle de 7,47 ps RMS selon des simulations faites avec le logiciel *Cadence*. Le CE a la flexibilité d'ajuster les temps d'étouffement et de recharge pour la PAMP tout en présentant une faible consommation de puissance ($\sim 0,33\text{ mW}$ à 33 Mcps). La conception du PAMP nécessite de supporter des tensions supérieures aux 3,3 V de la technologie. Pour répondre à ce problème, des transistors à drain étendu (DE-MOS) ont été réalisés. En raison de retards de production par les fabricants, les circuits n'ont pu être testés physiquement par des mesures. Les résultats de ce mémoire sont par conséquent basés sur des résultats de simulations avec le logiciel *Cadence*.

Mots-clés : Circuit d'étouffement, Photodiodes à avalanche monophotoniques (PAMP), Single Photon Avalanche Diode (SPAD), Intégration 3D hétérogène, Drain-Extended MOS (DEMOS), CMOS 130 nm 3D Tezzaron/Chartered, Tomographie d'émission par positrons (TEP)

À ma famille, mes amis et toute l'équipe du
GRAMS, qui n'ont jamais cessé de m'ap-
puyer.

REMERCIEMENTS

Après 4 ans et demi d'études pour le baccalauréat, combinés avec 3 ans au sein du Groupe de Recherche en Appareillage Médical de Sherbrooke (GRAMS) que je peux pratiquement appeler « famille », il est maintenant venu le temps de mettre un terme à cette grande étape de vie. Mon périple avec le GRAMS a commencé alors qu'un tout nouveau professeur à l'Université de Sherbrooke, Jean-François Pratte, est venu nous présenter ses travaux de recherche et sa vision de la nouvelle génération de scanners TEP. Ayant piqué ma curiosité, je n'ai pu m'empêcher d'aller discuter avec lui pour en savoir plus sur l'imagerie médicale. S'ensuivit alors une discussion d'environ 45 minutes, les premières d'une très longue série. Je veux donc tout d'abord remercier Jean-François, c'est grâce à lui si j'ai pu faire cette maîtrise. Je tiens à le remercier pour son énergie, sa créativité et son imagination, pour la patience qu'il a eue avec moi lorsque j'avais de la difficulté à saisir certains concepts, pour ses *cups Nespresso*, mais surtout pour son humour et plus que tout, d'avoir cru en moi. Je veux aussi remercier mon codirecteur Réjean Fontaine pour son côté terre à terre et rationnel, sans parler de son côté « zen ». Je me souviendrai toujours de sa règle de multiplier le temps par 2, grâce à laquelle j'estime beaucoup mieux mes échéanciers à présent.

Mes remerciements vont bien sûr également à mes collègues Marc-André Tétrault et Étienne Desaulniers Lamy (Ted) qui ont embarqué avec moi dans cette grande aventure de l'ASIC Tezzaron 3D, à travailler tôt et tard, sans compter leur aide si précieuse. Sans Marc-André, SVN m'aurait donné plus d'un cheveu gris, alors que Ted fournissait le support en musique, avec des chansons telles *Sweet Dreams* et *Voyage Voyage*. Je me dois également de remercier Caroline Paulin, Jonathan Bouchard, Luc Maurais et Sébastien Choquette pour leur grand support et leurs conseils pour la conception des cartes PCB. Merci à Audrey Corbeil Therrien pour ses muffins, à Benoit-Louis Bérubé pour toutes nos discussions sur les SPAD et pour son humour, ainsi qu'à toute l'équipe du GRAMS pour son support moral. Je tiens à remercier l'AGEG pour son apport en caféine et le Pub Willard qui a permis nombreuses grandes discussions et théories autour d'une douzaine d'ailes de poulet et d'une grosse Molson Dry avec Clamatto.

Évidemment, une grosse partie de mes remerciements revient à mes parents, mes sœurs et ma famille complète. Ils m'ont soutenu depuis le tout jeune âge, toujours encouragé à aller plus loin et me surpasser. Ils m'ont appuyé autant moralement que par le nombre incalculable de petits plats préparés avec amour. Je tiens à remercier mes amis Guillaume Houde et Simon St-Pierre pour leur écoute et leurs conseils. Finalement, un gros merci aux Foo Fighters, Metallica et ma guitare Godin qui m'ont permis de passer à travers les moments plus difficiles, ainsi qu'à Yoda de m'avoir donné la Force.

TABLE DES MATIÈRES

| | | |
|----------|---|-----------|
| 1 | INTRODUCTION | 1 |
| 2 | ÉTAT DE L'ART | 7 |
| 2.1 | Détecteurs | 7 |
| 2.1.1 | Tubes photomultiplicateurs | 7 |
| 2.1.2 | Photodiodes à avalanche en mode linéaire | 8 |
| 2.1.3 | Photodiodes à avalanche en mode Geiger | 9 |
| 2.1.4 | Photomultiplicateurs sur silicium | 12 |
| 2.2 | Circuits d'étouffement | 13 |
| 2.2.1 | Circuit d'étouffement passif | 14 |
| 2.2.2 | Circuit d'étouffement actif | 16 |
| 2.2.3 | Circuit d'étouffement mixte | 18 |
| 2.2.4 | Circuit d'étouffement avec comparateur de courant | 20 |
| 2.2.5 | Circuit d'étouffement à charge variable | 21 |
| 2.3 | Transistor à drain étendu | 23 |
| 2.4 | Intégration 3D | 26 |
| 2.4.1 | Historique | 27 |
| 2.4.2 | Interconnexions verticales | 30 |
| 2.5 | Modèles de simulation | 32 |
| 2.5.1 | Photodiodes à avalanche | 32 |
| 2.5.2 | Interconnexions verticales | 36 |
| 3 | ANALYSE ET CONCEPTION | 39 |
| 3.1 | Fonctionnement du circuit d'étouffement | 40 |
| 3.2 | Simulations | 42 |
| 3.2.1 | Modélisation de la PAMP | 42 |
| 3.2.2 | Modélisation du TSV | 45 |
| 3.2.3 | Électronique frontale 3,3V | 47 |
| 3.2.4 | Électronique frontale de haute tension | 50 |
| 3.2.5 | Bascule | 50 |
| 3.2.6 | Monostable du temps mort | 51 |
| 3.2.7 | Monostable du temps de recharge | 55 |
| 3.2.8 | Amorce et références de courant | 57 |
| 3.2.9 | Traducteur de niveau | 59 |
| 3.2.10 | Logique de contrôle | 60 |
| 3.2.11 | Générateur d'impulsions | 64 |
| 3.2.12 | Simulations de matrices | 66 |
| 3.2.13 | Schéma de la couche complète | 67 |
| 3.3 | Conception des masques | 67 |
| 3.3.1 | Masques du circuit et mise en matrice | 67 |
| 3.3.2 | Masques des DEMOS et des matrices à haute tension | 70 |

| | | |
|-------|---|-----|
| 3.3.3 | Circuits de tests | 71 |
| 3.3.4 | Plots d'interconnexions | 76 |
| 3.3.5 | Marques d'alignements et remplissage de métal | 78 |
| 3.4 | PCB et plan de tests | 80 |
| 3.4.1 | Fonctionnalité des circuits d'étouffement | 81 |
| 3.4.2 | Gigue des circuits d'étouffement | 84 |
| 3.4.3 | Courant de la cathode d'une PAMP | 85 |
| 3.4.4 | Mesure des temps des monostables | 87 |
| 3.4.5 | Caractérisation des transistors DEMOS | 87 |
| 3.4.6 | Tension de claquage des TSV | 88 |
| 4 | RÉSULTATS ET DISCUSSIONS | 89 |
| 5 | CONCLUSION | 93 |
| A | DIMENSIONS DES TRANSISTORS | 97 |
| B | VALEURS CONFIDENTIELLES | 99 |
| C | DIAGRAMMES D'ÉTATS | 101 |
| | LISTE DES RÉFÉRENCES | 105 |

LISTE DES FIGURES

| | | |
|------|--|----|
| 1.1 | Principe de base de la TEP | 3 |
| 2.1 | Fonctionnement du tube photomultiplicateur | 8 |
| 2.2 | Différence entre le mode linéaire et le mode Geiger | 9 |
| 2.3 | PAMP de procédé optoélectronique | 11 |
| 2.4 | PAMP de procédé CMOS standard | 11 |
| 2.5 | Schéma d'un SiPM | 12 |
| 2.6 | Photographie d'un SiPM | 12 |
| 2.7 | Circuit d'étouffement passif et modèle simple d'une PAMP | 15 |
| 2.8 | Circuit d'étouffement actif | 17 |
| 2.9 | Circuit d'étouffement mixte | 19 |
| 2.10 | Circuit d'étouffement avec comparateur de courant | 20 |
| 2.11 | Circuit de comparateur de courant | 20 |
| 2.12 | Circuit d'étouffement à charge variable | 21 |
| 2.13 | Schéma-bloc du circuit d'étouffement à charge variable | 21 |
| 2.14 | Exemple de transistors DEMOS | 24 |
| 2.15 | Lignes de champs équipotentiels de DEMOS | 25 |
| 2.16 | Les régions de canal, d'accumulation et de dérive dans un DEMOS | 25 |
| 2.17 | Nouvelle structure DEMOS avec le puits p^- isolé | 27 |
| 2.18 | Loi de Moore | 28 |
| 2.19 | Délais en fonction du procédé de fabrication | 28 |
| 2.20 | PAMP, circuit d'étouffement et électronique de logique | 29 |
| 2.21 | Matrice du circuit de la figure 2.20 | 29 |
| 2.22 | Système en intégration 3D | 30 |
| 2.23 | Modélisation simpliste d'une PAMP dans un PQC | 33 |
| 2.24 | Modélisation simpliste d'une PAMP avec 2 capacités parasites | 33 |
| 2.25 | Modélisation d'une PAMP incluant la capacité à l'anode, ainsi que les phénomènes d'automaintien et d'autoétouffement | 34 |
| 2.26 | Modélisation d'une PAMP avec branche directe et second claquage | 35 |
| 2.27 | Modélisation d'un TSV | 37 |
| 3.1 | Intégration 3D du projet développé | 40 |
| 3.2 | Schéma-bloc du circuit d'étouffement conçu | 41 |
| 3.3 | Schéma du circuit d'étouffement réalisé | 43 |
| 3.4 | Courbes des principaux signaux simulés avec le logiciel <i>Cadence</i> | 44 |
| 3.5 | Rayons d'un TSV | 45 |
| 3.6 | Schéma de l'électronique frontale | 48 |
| 3.7 | Déplacement des charges pour la décharge et la recharge | 49 |
| 3.8 | Schéma de la bascule | 51 |
| 3.9 | Schéma du monostable du temps mort | 53 |
| 3.10 | Schéma de la bascule de Schmitt | 54 |

| | |
|---|-----|
| 3.11 Schéma du monostable du temps de recharge | 56 |
| 3.12 Amorce : Schéma | 58 |
| 3.13 Amorce : Point d'équilibre | 58 |
| 3.14 Schéma de l'amorce et références de courant | 59 |
| 3.15 Schéma du traducteur de niveau | 60 |
| 3.16 Schéma-bloc de la logique de contrôle | 61 |
| 3.17 Schéma de la porte logique NON-OU | 61 |
| 3.18 Schéma de la porte logique ET | 62 |
| 3.19 Schéma de la porte logique NON-ET | 63 |
| 3.20 Schéma du générateur d'impulsions | 64 |
| 3.21 Comparaison entre les simulations avec la PAMP et avec le générateur d'impulsions | 65 |
| 3.22 Dessin des masques du circuit d'étouffement avec des transistors 3,3 V en entrée | 68 |
| 3.23 Dessin des masques d'une matrice de circuits d'étouffement | 70 |
| 3.24 Dessin des masques du circuit d'étouffement avec des transistors DEMOS en entrée | 71 |
| 3.25 Dessin des masques de la couche complète d'étouffement | 73 |
| 3.26 Dessin des masques d'un circuit d'étouffement de tests placé entre deux plots d'interconnexion | 74 |
| 3.27 Dessin des masques des monostables de tests | 74 |
| 3.28 Dessin des masques d'un transistor DEMOS | 75 |
| 3.29 Dessin des masques du premier test de TSV | 76 |
| 3.30 Dessin des masques du deuxième test de TSV | 77 |
| 3.31 Dessin des masques du troisième test de TSV | 77 |
| 3.32 Dessin des masques de la métallisation derrière la couche d'étouffement | 78 |
| 3.33 Dessin des masques de l'ouverture de la passivation pour les plots d'interconnexions | 79 |
| 3.34 Marques d'alignement | 80 |
| 3.35 Photo de la carte mère | 81 |
| 3.36 Photo de la carte fille | 81 |
| 3.37 Interface logicielle pour le contrôle des circuits de test | 82 |
| 3.38 Points de lectures des différents tests sur la carte fille | 83 |
| 3.39 Schéma électrique du point de test du courant de la cathode d'une PAMP | 85 |
| 3.40 Points de lectures des tests du courant de la cathode d'une PAMP | 86 |
| C.1 Diagramme d'états du contrôle de recharge de la PAMP. | 101 |
| C.2 Diagramme d'états du contrôle de détection/étouffement. | 102 |
| C.3 Diagramme d'états du contrôle de sortie par l'activateur. | 103 |
| C.4 Diagramme d'états du contrôle de la sortie du circuit d'étouffement. | 104 |

LISTE DES TABLEAUX

| | | |
|-----|---|----|
| 3.1 | Les 16 possibilités de temps morts simulés. | 52 |
| 3.2 | Les 16 possibilités de temps de recharge simulés. | 55 |
| 3.3 | Dimensions des transistors DEMOS. | 75 |

CHAPITRE 1

INTRODUCTION

Les technologies utilisant la détection photonique, c'est-à-dire la détection de lumière, sont en plein essor et prennent une place de plus en plus importante dans notre société. Elles s'appliquent à une variété de domaines, tels la cartographie, la météorologie, les recherches marines, le domaine minier, le domaine de l'automobile, l'imagerie médicale, l'exploration spatiale ou encore dans des appareils utilisés plus couramment comme les téléphones intelligents et les tablettes. Différents types d'imagerie permettent de développer ces applications. À ce titre, prenons comme premier exemple le LIDAR (*Light Detecting and Ranging*) [Northend *et al.*, 1966; Young, 2010] qui se base sur le même principe de fonctionnement que le RADAR conventionnel, mais utilise des impulsions lasers plutôt que des ondes radio [Schwarz, 2010]. Les impulsions heurtent l'objet et sont réfléchies à la source où des capteurs les détectent. En se basant sur la vitesse de la lumière et le temps de vol de l'impulsion, soit le temps que cette lumière a mis pour faire l'aller-retour entre la source et l'objet, il est possible de connaître la distance entre ceux-ci. Ce principe a d'ailleurs été utilisé pour développer la voiture *MadeInGermany* par le groupe de recherche *AutoNOMOS Labs*, dont la conduite est automatisée et autonome [AutoNOMOS Labs, 2011]. La voiture utilise, entre autres, un module rotatif de la compagnie Velodyne Lidar, Inc. qui comporte plusieurs émetteurs et récepteurs, permettant une lecture sur 360° et ainsi de recréer un environnement en trois dimensions autour du véhicule [Velodyne Lidar, Inc., 2010]. Un deuxième exemple pourrait être l'exploration spatiale où l'on utilise la détection de radiations Cherenkov comme type d'imagerie. Ce type de radiation survient lorsqu'une particule chargée se déplace plus rapidement que la vitesse de la lumière dans un milieu diélectrique et perturbe au passage la polarisation des couches électroniques des atomes du milieu, créant une émission radiative [Mather, 1951]. La détection de ces photons est particulièrement utile pour la sécurité des réacteurs nucléaires. Ce type d'imagerie trouve également son intérêt pour la détection de rayons gamma provenant de l'espace pour les études astrophysiques, ou encore dans la détection de neutrino [Los Alamos National Laboratory, 2012; Pierre Auger Observatory, 2012; The Imagine Team, 2011].

Le domaine médical bénéficie grandement des technologies de détections photoniques et de l'imagerie 3D. La tomographie optique diffuse (TOD) en est un exemple. Elle utilise la

lumière avec des longueurs d'onde du rouge et des infrarouges pour caractériser en trois dimensions des tissus biologiques en profondeur [Boffety, 2010; Dubois, 2007]. En utilisant des radiations non ionisantes, elle se voit particulièrement inoffensive comparativement à d'autres types d'imagerie médicale [Boffety, 2010; Nouizi, 2012]. En contrepartie, la diffusion et l'absorption de la lumière dans le patient rendent difficile une bonne lecture pour des milieux plus épais. C'est pourquoi la TOD est utile dans le dépistage pour le cancer du sein ou encore dans l'étude de l'activité cérébrale, mais devient moins performante pour la détection de tumeurs cancéreuses à l'intérieur d'un patient. Une des techniques d'imagerie les plus utilisées dans ce cas est la tomographie d'émission par positrons (TEP).

Le Groupe de Recherche en Appareillage Médical de Sherbrooke (GRAMS) se spécialise dans le domaine des scanners TEP/TDM et prochainement TOD. Ce groupe de recherche vise à sans cesse améliorer ce type d'imagerie médicale. C'est par ailleurs au sein de ce groupe de recherche que le présent projet de maîtrise a pris forme. La TEP est, entre autres, utilisée afin d'observer l'activité métabolique des organes ou de cellules spécifiques. Elle est souvent combinée à de l'imagerie de rayons X, afin de bien localiser l'emplacement de l'activité métabolique dans le corps du patient. La TEP utilise des marqueurs radioactifs injectés au patient sous analyse. Le principal marqueur utilisé en oncologie est un analogue du glucose dont les cellules à hautes activités métaboliques, telles que les cellules cancéreuses, consomment davantage que les tissus sains. Le radioisotope inclus dans le traceur se désintègre par l'émission d'un positron (un électron de charge positive). Ce dernier s'annihile ensuite avec un électron du milieu, résultant en l'émission de deux photons de 511 keV à 180° l'un de l'autre. Ces photons d'annihilation sont détectés par un anneau de scintillateurs et de capteurs entourant le patient (figure 1.1). Les scintillateurs sont principalement des cristaux composés d'orthosilicates couplés à des matériaux lourds comme le lutécium, l'yttrium ou le gadolinium, dopés au cérium, comme dans le cas des LYSO ou LGSO [Cherry *et al.*, 2003; Tetrault *et al.*, 2008]. Leur pouvoir d'arrêt permet d'absorber les photons de 511 keV sur approximativement 1 cm. Ces scintillateurs transforment l'énergie des photons d'annihilation en une pluie de photons du côté bleu du spectre lumineux (420 nm) [Cherry *et al.*, 2003; Fontaine *et al.*, 2009]. Des photodiodes à avalanche ou des tubes photomultiplicateurs changent à leur tour cette lumière en une douche d'électrons amplifiée et traitée par la caméra TEP [Fontaine *et al.*, 2003]. En connaissant l'endroit d'arrivée de ces deux photons d'annihilation dans l'anneau, il est possible de tracer une ligne ou un total de réponse entre les deux détecteurs où l'annihilation a eu lieu. La combinaison de toutes les lignes ou tubes de réponse permet de faire la reconstruction d'image à l'intérieur du patient [Cherry *et al.*, 2003; Fontaine *et al.*, 2003].

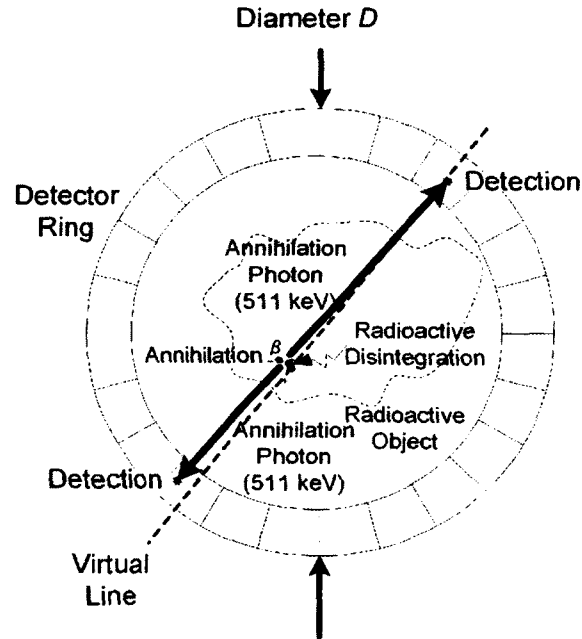


Figure 1.1 Principe de base de la TEP [Fontaine, 2010].

Tout comme pour la technologie du LIDAR, la TEP peut utiliser le temps de vol (*Time of Flight*, ToF) dans les scanners cliniques pour les humains afin permet de circonscrire l'annihilation sur la ligne de réponse et ainsi améliorer le contraste de l'image [Cherry *et al.*, 2003]. Plus la gigue temporelle de la mesure sur la paire scintillateur-détecteur sera faible, meilleure sera la précision sur le temps de vol des photons d'annihilation et meilleur sera le contraste sur l'image. Afin d'optimiser le rendement de ce type d'imagerie, en plus de chercher à avoir les détecteurs les plus précis possible sur la mesure temporelle, il faut que les détecteurs couvrent le plus grand angle solide (i.e. le pourcentage de la surface sous les scintillateurs). De cette façon, il y a plus de probabilités de détecter les premiers photons provenant des scintillateurs, lesquels contiennent le plus d'informations utiles et pouvant améliorer la mesure du temps de vol.

Ainsi, que ce soit pour la technologie du LIDAR, des radiations Cherenkov, de la TOD ou encore pour la TEP, les enjeux technologiques restent sensiblement les mêmes. Afin d'obtenir les meilleures performances pour chacune de ces applications, les détecteurs doivent avoir une forte probabilité de détection de photons et une faible gigue temporelle, tout en ayant une faible consommation de puissance. Dans la même optique, la surface utilisée par les détecteurs doit être la plus large possible, afin d'avoir le plus grand ratio de surface photosensible. Des travaux de recherche s'orientent vers le développement de nouvelles ap-

proches technologiques permettant d'obtenir de meilleurs compromis en considérant tous ces aspects.

Dans le contexte de la conception d'une troisième génération de ses scanners dans lequel s'inscrit la présente étude, l'approche adoptée par l'équipe du GRAMS est d'utiliser la photodiode à avalanche monophotonique (PAMP) matricée sous les scintillateurs pour atteindre un autre niveau de performance. Dans la nouvelle approche, chaque PAMP serait couplée individuellement à un circuit d'étouffement sur une seconde couche sous les photodiodes grâce à l'intégration 3D. En plus d'étouffer et de recharger chaque PAMP, le circuit d'étouffement contrôlerait son temps mort afin de réduire son bruit postimpulsionnel. Chaque circuit d'étouffement communiquerait avec une troisième couche constituée de l'électronique de traitement numérique de signaux. À la lumière de ce contexte, il en découle la question de recherche suivante :

« Est-il possible d'étouffer une photodiode à avalanche en mode Geiger avec un circuit d'étouffement en technologie CMOS sous-micrométrique contraint en espace à la taille de la photodiode dans une architecture 3D, et ce, le plus rapidement possible, avec une gigue temporelle minimale, tout en contrôlant son temps mort et en limitant la consommation de puissance, pour un module de photodétection avec résolution monophotonique ? »

À partir de la question de recherche s'ensuivent les objectifs de ce mémoire liés à celle-ci. L'objectif général du mémoire est de concevoir et d'intégrer un circuit d'étouffement pour une photodiode à avalanche en mode Geiger pour l'architecture 3D d'un nouveau scanner TEP. De là suivent plusieurs objectifs spécifiques. En utilisant le procédé de fabrication de Tezzaron/Chartered de CMOS 130 nm 3D, le premier objectif spécifique vise à utiliser un espace de $50 \times 50 \mu\text{m}^2$ pour le circuit d'étouffement, soit le même espace que pour les photodiodes de la couche au-dessus. De cette façon, le circuit pourra être répété sous la matrice de PAMP, elle-même sous un scintillateur de $1,2 \times 1,2 \text{ mm}^2$. Le second objectif spécifique s'adresse au contrôle du temps mort de la photodiode. Le circuit d'étouffement doit offrir la possibilité de différentes durées de temps mort à partir de bits de configuration provenant de la couche de l'électronique de traitement numérique de signaux. Le troisième objectif spécifique exige que le circuit d'étouffement tolère la plus grande tension d'excès possible avec la technologie CMOS 130 nm 3D de Tezzaron/Chartered. Le quatrième et dernier objectif spécifique demande que le circuit ait une faible consommation de puissance.

Dans la perspective d'atteindre ces objectifs, une revue de la littérature sera premièrement présentée dans ce document au chapitre 2. Elle portera sur l'étude de différents photodétecteurs à la section 2.1, ainsi que sur une variété de circuits d'étouffement (section 2.2)

nécessaires aux photodiodes à avalanche monophotoniques. Les transistors à drain étendu auront aussi droit à une part de chapitre en section 2.3, cette forme de transistor pouvant supporter une plus grande tension que les transistors conventionnels présentera un avantage certain pour le contrôle de photodétecteurs. La revue de littérature fera également mention de l'intégration 3D en section 2.4. La section 2.5 présentera les modèles de simulation nécessaires à la conception du circuit d'étouffement. Cette étude a permis de modéliser analytiquement par simulations, avec le logiciel *Cadence*, les circuits d'étouffement les plus prometteurs et de développer, ainsi qu'optimiser, le circuit répondant le mieux aux exigences. Le troisième chapitre présentera ce circuit réalisé à l'intérieur de ce projet de maîtrise. Ce chapitre portera sur le fonctionnement global du circuit d'étouffement conçu en section 3.1, alors que la section 3.2 concernera les simulations et fera la présentation de chacun des blocs du circuit. La section 3.3 parlera du dessin des masques permettant l'envoi de fichiers GDS en fonderie, afin que Chartered et Tezzaron Semiconductor fabriquent la puce. Les résultats obtenus et des discussions s'en suivront au chapitre 4, afin de terminer avec une conclusion sur ce mémoire.

CHAPITRE 2

ÉTAT DE L'ART

Les applications utilisant une détection avec une résolution monophotonique s'évaluent selon trois critères de performance : une bonne probabilité de détection de photons, une gigue temporelle minimale et une faible consommation de puissance. Ces critères seront abordés dans les cinq sections de ce chapitre. La première section présente différents types de détecteurs photoniques. La seconde section porte sur divers types de circuit d'étouffement nécessaires aux photodiodes à avalanche monophotoniques (PAMP). La section 2.3 introduit les transistors à drain étendu (DEMOS) permettant de supporter une tension plus élevée que les transistors standards. L'intégration 3D est ensuite présentée à la section 2.4, avant de terminer avec les modèles de simulation nécessaires à l'élaboration d'un circuit d'étouffement en cinquième section.

2.1 Détecteurs

Le type de détecteurs utilisé dans une application de détection avec une résolution monophotonique est l'élément clé pour obtenir de bonnes performances. Il doit être compact et présenter la probabilité optimale de détection de photons, tout en ne dégradant pas les deux autres critères de performance. Plusieurs types de détecteurs sont présents dans la littérature : les tubes photomultiplicateurs [Cherry *et al.*, 2003], les photodiodes à avalanche [Zappa *et al.*, 2007], les photomultiplicateurs sur silicium [Eraerds *et al.*, 2007] et les photodiodes à avalanche opérées en mode Geiger [Stapels *et al.*, 2006; Stoppa *et al.*, 2009]. Une revue de tous ces différents types de détecteurs est au cœur de cette première section.

2.1.1 Tubes photomultiplicateurs

Le premier type de détecteur est le tube photomultiplicateur (TPM). Son principe de fonctionnement se base sur une série de dynodes à l'intérieur de tubes à vide qui aura pour effet de multiplier les charges le traversant (figure 2.1). Le principe d'amplification débute par l'utilisation d'une photocathode, partie photosensible du tube, d'où des électrons sont libérés [Cherry *et al.*, 2003]. Ces électrons sont accélérés par le champ électrique appliqué entre les dynodes. Le champ électrique inter-dynodes est assez grand pour que les

électrons acquièrent une énergie suffisante pour arracher chacun quelques électrons sur les dynodes subséquentes. À la sortie du TPM, toutes les charges décrochées des dynodes sont recueillies par l'anode et créent l'impulsion recherchée permettant de détecter l'arrivée du photon. Le TPM ayant un gain fixe pour le nombre de photons absorbés, l'amplitude de ce signal de sortie se trouve linéaire avec le flux lumineux incident. Cependant, ces tubes prennent beaucoup d'espace, sont coûteux et fragiles. De plus, ils nécessitent une polarisation de haute tension, soit entre 1000 V et 2500 V typiquement, et sont également très sensibles aux champs magnétiques.

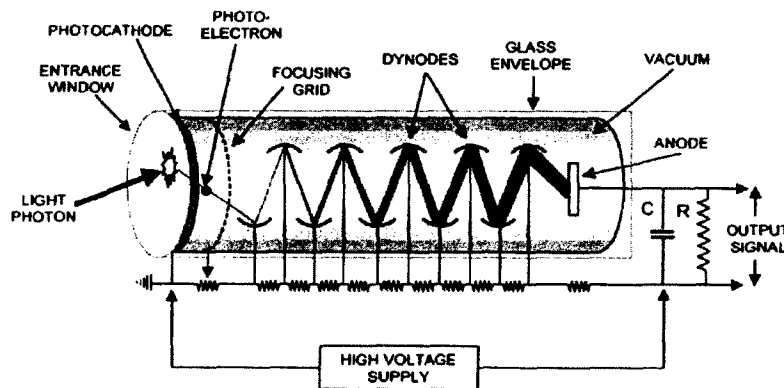


Figure 2.1 Fonctionnement du tube photomultiplicateur [Cherry *et al.*, 2003].

2.1.2 Photodiodes à avalanche en mode linéaire

Devant trouver des photodétecteurs plus compacts pour répondre aux besoins spécifiques de la TEP sur petits animaux, les chercheurs se sont tournés vers la photodiode à avalanche (PDA). Lorsqu'elle est utilisée en mode linéaire, c'est-à-dire avec une tension en inverse assez élevée, mais en deçà de la tension de claquage, elle peut être utilisée pour obtenir une réponse linéaire en fonction du flux lumineux incident. Lorsqu'un photon est absorbé dans sa zone de déplétion, il crée une paire électron-trou. Les porteurs négatifs photogénérés, accélérés par le champ électrique appliqué aux bornes de la photodiode, créent d'autres paires électron-trou par ionisation par impact qui vont à leur tour en former d'autres. Le courant d'avalanche résultant possède un gain fixe selon les caractéristiques de la photodiode. Pour une paire électron-trou créée par un photon incident, une multiplication de paires variant de quelques dizaines à quelques centaines est obtenue. Comme les PDA produisent un signal relativement faible et sont bruyantes, ce signal doit être amplifié par un préamplificateur de charges avec un très faible bruit afin de ne pas détériorer son ratio

signal sur bruit (*Signal-to-Noise Ratio*, SNR). Ce photodétecteur présente deux désavantages quant à la gigue temporelle ; le premier provient du fait que la PDA requiert quelques photons pour que son avalanche soit détectable, posant une incertitude sur le temps d'arrivée du premier photon, en considérant que l'arrivée des photons suit une distribution de Poisson. Le second provient du faible rapport signal sur bruit de la photodiode accentuant cette incertitude.

2.1.3 Photodiodes à avalanche en mode Geiger

Les photodiodes à avalanche offrent une seconde option puisqu'elles peuvent être conçues pour opérer plus particulièrement au-delà de la tension de claquage. Lorsqu'elles sont polarisées dans cette condition, elles portent la dénomination de photodiodes à avalanche monophotoniques (PAMP) opérées en mode Geiger, plus communément appelées selon l'acronyme anglais SPAD (*Single Photon Avalanche Diode*) [Stapels *et al.*, 2006; Stoppa *et al.*, 2009]. Le fort champ électrique appliqué entre ses bornes amène la photodiode dans un état métastable où un seul photon absorbé dans la région photosensible suffit à déclencher une avalanche divergente. Cette dernière provient de l'ionisation par impact impliquant autant les porteurs positifs que négatifs résultant de l'énorme potentiel appliqué à la PAMP (figure 2.2). Il en résulte ainsi un gain pratiquement infini où seule la résistance interne de la photodiode limite le courant [Zappa *et al.*, 2007]. Une fois déclenchée, l'avalanche s'emballe et cause l'autodestruction de la photodiode si cette dernière n'est pas rapidement étouffée par un circuit capable de diminuer sa tension sous la tension de claquage. Par sa vitesse de réaction, son emballement dans l'avalanche et par le fait qu'elle nécessite qu'un seul photon pour s'activer, la photodiode à avalanche en mode Geiger est utilisée plutôt comme un déclencheur, contrairement à son rôle d'amplificateur lorsqu'elle est utilisée en mode linéaire.

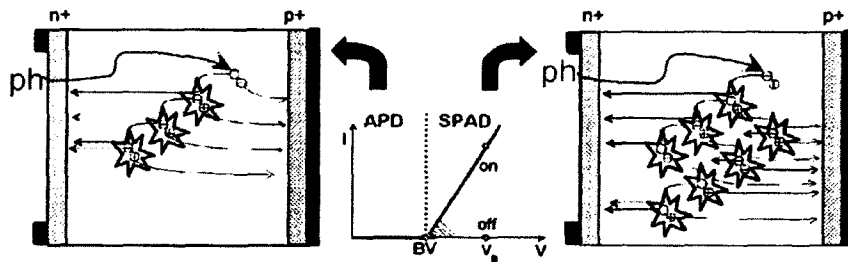


Figure 2.2 Différence entre une photodiode à avalanche en mode linéaire (à gauche) et une photodiode à avalanche en mode Geiger (à droite) [Zappa *et al.*, 2007].

Tout comme les PDA opérées en mode linéaire, les PAMP possèdent certains bruits influençant leurs performances. Quatre types s'y retrouvent : thermique, à effet tunnel, postimpulsionnel et diaphonie optique. Chacune de ces sources de bruit sera abordée. L'excitation thermique dans la photodiode peut générer des électrons libres et parfois déclencher des avalanches appelées bruit thermogénéré, même en l'absence de lumière [Stipcevic, 2009]. Le taux de déclenchements issu de ce bruit dans le noir est plus connu sous la terminologie de bruit d'obscurité ou sous l'abréviation DCR (de l'anglais *Dark Count Rate*). Ce bruit constitue l'élément principal du bruit interne du photodétecteur où le diamètre de la région photosensible influence proportionnellement le DCR [Zappa *et al.*, 2007]. Pour sa part, le bruit à effet tunnel provient du déclenchement d'une avalanche dû au passage direct d'un électron de la bande de valence à la bande de conduction en raison du fort champ électrique de la PAMP. Ce type de bruit est également inclus dans la mesure du DCR. Dans le bruit postimpulsionnel, des porteurs sont attrapés après une avalanche dans des pièges profonds autour d'atomes de matière, issus d'imperfections dans la structure cristalline. Le porteur se délivre alors du piège, déclenchant une seconde avalanche corrélée en temps avec l'avalanche précédente [Rech *et al.*, 2009]. Une méthode utilisée pour réduire l'impact causé par ce bruit consiste à garder la photodiode étouffée pendant un certain temps. De cette façon, les porteurs ont le temps de se libérer de leur piège sans causer une seconde avalanche, puisque la photodiode n'est toujours pas polarisée au-delà de sa tension de claquage. Finalement, la diaphonie optique provient de photons libérés par une photodiode voisine pendant son avalanche. Cette situation apparaît lorsqu'un porteur ayant une grande vitesse qui interagit avec un noyau atomique. Dans cette interaction, le porteur est dévié de sa trajectoire et grandement ralenti provoquant l'émission d'un photon, effet connu sous le nom *Bremsstrahlung* [Cherry *et al.*, 2003]. La température joue aussi un rôle dans le rendement des PAMP. Plus la température est élevée, plus le bruit thermogénéré est de grande envergure. En contrepartie, le bruit postimpulsionnel diminue. En effet, à de plus hautes températures, le temps de vie des pièges profonds diminue. De ce fait, la probabilité de déclencher une avalanche après un même temps mort diminue [Ghioni *et al.*, 2007]. Il y a donc un compromis à faire entre les deux et une température d'opération optimale à atteindre. De plus, la tension de seuil pour le claquage de la photodiode augmente avec la température. Subséquemment, pour une même tension aux bornes de la photodiode, plus la température augmente, plus la tension d'excès au claquage diminue, diminuant ainsi la probabilité de déclencher une avalanche et par le fait même l'efficacité de détection de photons (*Photon-Detection Efficiency*, PDE).

Les photodiodes peuvent être fabriquées dans divers procédés comme le procédé opto-électronique ou le procédé CMOS standard, telles que présentées aux figures 2.3 et 2.4

respectivement. Chacun présente ses avantages et ses inconvénients. La PAMP fabriquée en procédé optoélectronique présente un DCR et un bruit postimpulsionnel plus faible que celui d'une PAMP en procédé CMOS. En effet, la possibilité de changer les profils de dopant et de mieux contrôler la qualité du matériau du photodétecteur permet entre autres de réduire le nombre de pièges profonds de la zone d'avalanche et ainsi réduire le bruit. En ayant le contrôle total sur les étapes du procédé de fabrication, le concepteur a la flexibilité d'optimiser le PDE de la photodiode, son DCR ou encore sa résolution en temps. Cette latitude est impossible en procédé CMOS puisqu'il n'offre aucune flexibilité au concepteur quant aux étapes de fabrication. En revanche, les photodiodes de procédé CMOS possèdent un coût de production inférieur au procédé optoélectronique dû à sa production de masse. La grande disponibilité de fabrication des photodiodes en procédé CMOS grâce aux gaufres à projets multiples (*Multi-Project Wafer*, MPW) leur confère un autre grand avantage. De plus, le procédé CMOS permet de localiser l'électronique de lecture de la photodiode juste à côté de cette dernière sur la même pièce, une technique difficile en procédé optoélectronique. Une autre grande différence entre les deux types de photodiodes se trouve au niveau de leur tension d'alimentation. La photodiode de procédé optoélectronique possède une tension de claquage plus élevée que celle de procédé CMOS standard. Sa tension de polarisation doit donc être plus élevée à ses bornes pour obtenir une même tension d'excès.

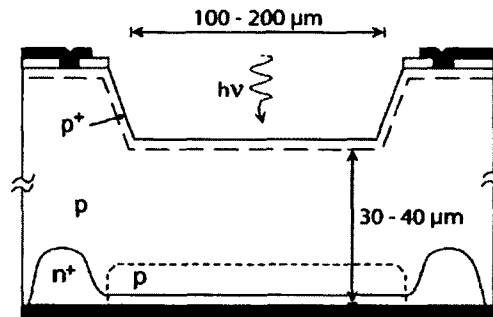


Figure 2.3 PAMP de procédé optoélectronique [Cova *et al.*, 2004].

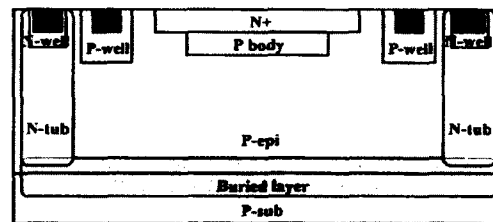


Figure 2.4 PAMP de procédé CMOS standard [Kim *et al.*, 2008].

2.1.4 Photomultiplicateurs sur silicium

Les PAMP étant compactes et facilement matricées, elles ont permis une architecture particulière de détecteurs, soit les photomultiplicateurs sur silicium, plus communément appelés selon le terme anglais *Silicon Photomultipliers* (SiPM) [Eraerds *et al.*, 2007]. Ces derniers consistent en des matrices de PAMP, dont chacune des photodiodes est étouffée passivement (section 2.2.1) par une résistance (figure 2.5). Les paires de photodiodes et résistances d'étouffement sont connectées en parallèle les unes aux autres et le signal de sortie représente la somme analogique des photodiodes ayant fait feu. Les SiPM servent ainsi à compter le nombre de photons absorbés par surface totale, comme sur la photographie de la figure 2.6, plutôt que par PAMP individuelle.

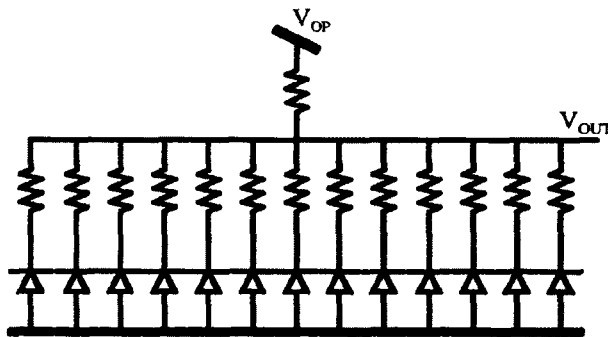


Figure 2.5 Schéma d'un SiPM [Charbon, 2008].

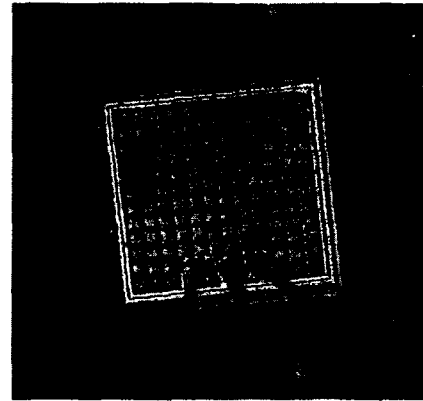


Figure 2.6 Photographie d'un SiPM [Eraerds *et al.*, 2007].

Malgré que les SiPM soient de conception simple, ils présentent certains désavantages comparativement à une lecture de PAMP individuelle. En premier lieu, la lecture du signal de sortie pour connaître le nombre de PAMP ayant été déclenchées repose sur le niveau de courant sommé. Afin que la lecture soit précise quant au nombre de photodiodes ayant fait feu, il faut que chaque PAMP du SiPM possède les mêmes caractéristiques et fournisse la même quantité de courant. Cependant, les défauts et différences de fabrication d'une photodiode à l'autre ne peuvent garantir cette constance. De plus, les changements de température faisant varier la tension de claquage des photodiodes, la tension d'excès appliquée aux PAMP est non-constante, tout comme le courant d'avalanche en résultant. En second lieu, la résolution en temps est limitée due à la différence de distance séparant chaque photodiode du nœud de lecture. En effet, une variation est due au temps de propagation qui sera plus long pour une PAMP se situant au bout de la chaîne du SiPM.

que pour une autre collée sur le nœud de lecture. De plus, les SiPM souffrent des mêmes désavantages que le circuit d'étouffement passif, qui seront présentés à la section 2.2.1.

En comparant ces différents photodétecteurs entre eux, soient les TPM, PDA, PAMP et SiPM, les PAMP semblent les plus prometteurs quant à la sensibilité aux premiers photons et à la gigue temporelle. C'est pourquoi le GRAMS a choisi d'utiliser ces photodétecteurs pour leur prochain détecteur TEP.

2.2 Circuits d'étouffement

Tel que mentionné précédemment, la photodiode à avalanche nécessite un circuit d'étouffement lorsqu'elle est opérée en mode Geiger. Selon [Tisa *et al.*, 2007; Zappa *et al.*, 2007], le circuit d'étouffement vise quatre principaux buts :

- détecter l'augmentation du courant avalanche ;
- générer à la sortie une impulsion synchrone avec le changement de courant ;
- étouffer l'avalanche en abaissant la polarisation en deçà de la tension de claquage ;
- remettre la polarisation au-delà de la tension de claquage.

Outre ces principaux objectifs, il faut prendre en considération d'autres caractéristiques, afin d'optimiser le rendement du circuit d'étouffement. Tout d'abord, une bonne conception du circuit d'étouffement est primordiale, car elle influence directement la gigue temporelle obtenue de la paire que ce circuit forme avec la PAMP. Si le but recherché est d'utiliser les photodiodes comme détecteurs matricés, il faut que le circuit soit de plus petite taille possible pour que la majorité de l'espace soit photosensible. Afin de réduire la consommation de puissance, la dissipation de chaleur, la diaphonie optique et le nombre de charges prises dans des pièges profonds, le nombre total de charges de l'avalanche doit être réduit à sa plus petite valeur [Gallivanoni *et al.*, 2010]. Également, la recharge doit se faire dans les plus brefs délais pour que la photodiode soit toujours dans les conditions optimales avant de déclencher une avalanche. Dépendamment de l'architecture de la photodiode, les électrodes du dispositif ne présentent pas la même capacité parasite. Pour une photodiode de type n^+p , la terminaison la moins capacitive est la cathode, alors que dans le cas d'un type p^+n , il s'agit de l'anode [Gallivanoni *et al.*, 2010]. En effet, la section servant de puits à la photodiode offre une plus grande aire capacitive avec le substrat, l'électrode y étant connectée étant donc la plus capacitive des deux [Tisa *et al.*, 2008b]. C'est pourquoi le circuit d'étouffement n'est pas branché à la même électrode d'un type de photodétecteurs à

un autre. Relié au côté le moins capacitif de la photodiode, l'éteuffement de la photodiode se fait plus rapidement, le nombre de charges émis par l'avalanche circulant dans la zone de déplétion est moins élevé et la recharge se fait, elle aussi, plus rapidement, réduisant les chances que la photodiode se déclenche avant d'avoir atteint la tension de polarisation désirée. Avec un nombre de charges réduit, moins de porteurs se prennent dans des pièges profonds, diminuant ainsi le bruit postimpulsionnel. Si le circuit d'éteuffement permet de contrôler le temps mort de la photodiode, c'est-à-dire la période de temps où la PAMP est gardée éteuffée intentionnellement, les porteurs pris dans les pièges profonds ont une probabilité plus élevée d'être relâchés avant que la photodiode ne soit polarisée au-dessus de sa tension de claquage. Le bruit postimpulsionnel se retrouve de cette façon encore diminué [Zappa *et al.*, 2007].

À la lumière de ces différents objectifs, il devient plus facile de comparer les avantages et inconvénients des différentes architectures de circuits d'éteuffement entre elles. Il en existe différents types, pouvant avoir un éteuffement passif ou actif, ainsi qu'une recharge passive ou active également. Parmi les diverses structures de circuits d'éteuffement se retrouvent principalement le circuit d'éteuffement passif, le circuit d'éteuffement actif, le circuit d'éteuffement mixte, le circuit d'éteuffement avec comparateur de courant et le circuit d'éteuffement à charge variable.

2.2.1 Circuit d'éteuffement passif

Le premier type, et de surcroît le plus simple, est le circuit d'éteuffement passif (*Passive Quenching Circuit*, PQC). Il n'est constitué que de deux résistances et d'un comparateur de tension, ne nécessitant donc que très peu de composants. Il se compose d'une résistance à haute impédance (R_L) pour absorber la majorité de la tension d'éteuffement et une autre de plus faible impédance (R_S) pour l'entrée du comparateur de tension (figure 2.7).

Afin de mieux comprendre le fonctionnement de l'éteuffement et de la recharge de la PAMP, un modèle équivalent simplifié de la photodiode à avalanche est présenté à la droite du schéma (figure 2.7). D'autres modèles plus complexes seront présentés ultérieurement dans ce document (section 2.5.1). À la figure 2.7, V_B représente la tension de claquage de la photodiode, R_D représente la résistance interne, alors que C_D équivaut à la capacité de jonction de la diode. Le condensateur C_P ajoute la capacité parasite vue à la cathode par la résistance R_L . Dans le but d'opérer la PAMP en inverse, la cathode de la photodiode est branchée du côté positif de l'alimentation, alors que l'anode est connectée à la masse par l'entremise du circuit d'éteuffement. La branche de la résistance R_D étant ouverte lorsque la tension d'alimentation V_A est augmentée, aucun chemin direct n'est disponible pour le

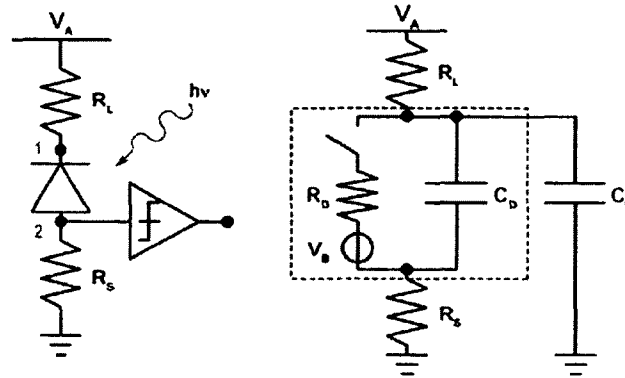


Figure 2.7 À gauche : Circuit d'éteuffement passif. À droite : Circuit d'éteuffement passif avec une modélisation simple d'une PAMP [Tisa *et al.*, 2007].

courant pour se rendre à la mise à la terre, chargeant ainsi la capacité de jonction et la capacité parasite. Lorsque la tension de claquage est excédée, un seul photon créant une paire électron-trou suffit pour déclencher l'avalanche. Cela est représenté par la fermeture du commutateur, permettant ainsi au courant de circuler librement à travers les résistances R_L , R_D et R_S jusqu'à la masse, en plus des charges accumulées aux capacités C_D et C_P qui veulent se décharger. La résistance à haute impédance R_L développe une tension à ses bornes à l'aide du courant qui la traverse, ce qui diminue la tension aux bornes de la photodiode, et ce, jusqu'à descendre en deçà de la tension de claquage de la photodiode [Cova *et al.*, 1996]. L'avalanche ne peut être soutenue et le courant cesse de traverser la diode, équivalant à rouvrir l'interrupteur. La résistance de plus faible impédance participe aussi à cette tâche, mais à plus petite échelle. Son rôle consiste plutôt à développer une tension à ses bornes assez grande pour dépasser la tension de seuil et déclencher le comparateur de tension, avisant ainsi la logique du système de l'événement [Tisa *et al.*, 2007]. Une fois la photodiode étouffée et aucun courant ne circulant par la branche R_D , la tension aux bornes des résistances redevient nulle, repolarisant la photodiode au-delà de sa tension de claquage et la réinitialisant dans ses conditions optimales, prête à se décharger de nouveau [Niclass *et al.*, 2005].

Malgré que ce type de circuit d'éteuffement soit simple et ne comporte que peu de composants, il n'en demeure pas moins que le choix de ceux-ci doit être fait judicieusement. La valeur de la résistance R_S doit être assez élevée pour que la tension développée à ses bornes, grâce au courant la traversant, garantisse une détection par le comparateur. En prime, plus la résistance est de grande taille, plus elle aide à l'éteuffement rapide de la photodiode, ainsi qu'à réduire le nombre de charges libérées, affectant par le fait même le bruit de la diode. En contrepartie, l'enjeu est de s'assurer que la tension formée à ses

bornes ne soit pas trop élevée, afin de protéger l'électronique du comparateur branché à ce nœud. Le choix de la résistance R_L est quant à lui plus critique, puisqu'il doit faire un compromis entre la constante de temps d'éteuffement et celle de repolarisation. En effet, si la résistance est élevée, la constante de temps d'éteuffement est plus courte et le nombre de charges libérées est réduit, diminuant conséquemment le bruit et la puissance dissipée. Lorsque la résistance est suffisamment élevée, son influence sur la constante de temps d'éteuffement devient limitée par le temps de transit des porteurs dans la jonction interne de la photodiode, représenté par la constante du circuit RC interne de la figure 2.7 [Mita et Palumbo, 2008]. En revanche, la recharge de la photodiode doit se faire à travers cette même résistance, ce qui ralentit le retour à l'équilibre aux bornes de R_L étant donné sa valeur élevée. Une recharge lente de la photodiode amène plusieurs problèmes. Tout d'abord, cela augmente la probabilité qu'un photon redéclenche la photodiode avant qu'elle ait atteint ses conditions optimales d'opération, soit la tension excédentaire à la tension de claquage désirée. Ce fait étant, les conditions varient d'un déclenchement à l'autre et manquent de constance. Dans certaines architectures, la résistance R_L se situe entre la photodiode et la résistance R_S . La recharge étant lente, la photodiode peut refaire feu avant même que la tension ne soit descendue assez pour être sous la tension de seuil du comparateur. Le système ne détecte alors pas ce photon.

Cependant, ce circuit possède plusieurs avantages. Ses principaux atouts sont sa simplicité de conception et son faible nombre de composants. En utilisant des composants passifs, ce type de circuit d'éteuffement commence à agir dès le déclenchement de l'avalanche comparativement à d'autres types de circuit qui peuvent nécessiter un certain délai. À son désavantage, les temps d'éteuffement et de recharge sont lents. L'efficacité de détection de photons pour une application de détection monophotonique est donc plus faible, puisque le circuit prend beaucoup de temps à réinitialiser la photodiode prête à recevoir un nouveau photon. Finalement, le temps mort de la photodiode n'est pas contrôlable avec cette architecture. Le circuit est de ce fait plus susceptible de redéclencher, dû à un porteur se libérant d'un piège profond dans la photodiode, étant donné que le bruit postimpulsionnel est plus élevé dans les instants suivant une avalanche [Tisa *et al.*, 2008b].

2.2.2 Circuit d'éteuffement actif

Afin de parer aux désavantages du circuit d'éteuffement passif, le circuit d'éteuffement actif a fait son apparition (*Active Quenching Circuit*, AQC) [Tisa *et al.*, 2007]. Son électronique étant plus étoffée, sa conception est plus complexe que celle du circuit passif. En effet, il est constitué d'un transistor de type PMOS pour l'éteuffement de la photodiode, d'un

transistor de type NMOS pour la réinitialisation et d'un comparateur de tension pour la détection d'un événement, en plus de toute la logique de contrôle (figure 2.8).

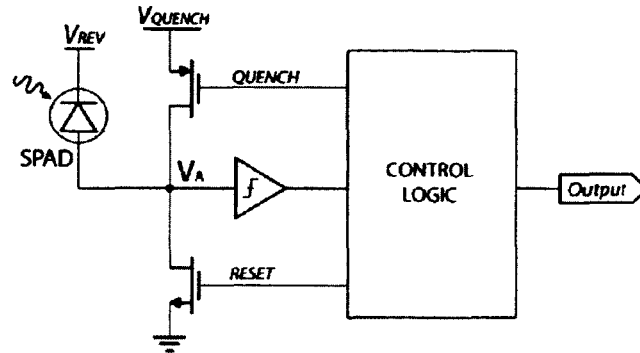


Figure 2.8 Circuit d'éteuffement actif [Tisa *et al.*, 2008a].

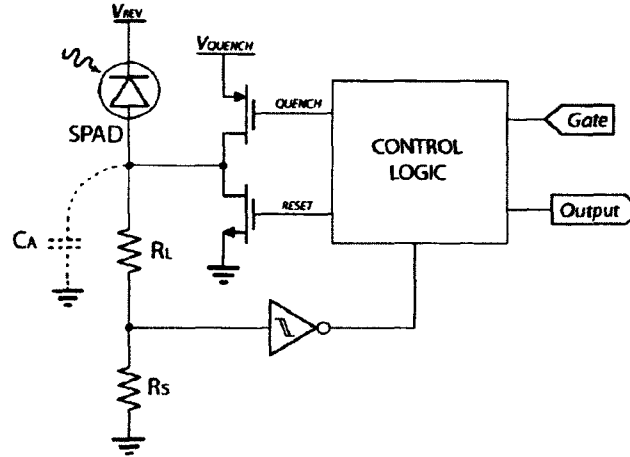
À l'état initial, la tension V_A à l'anode de la photodiode est de 0 V. Toute la tension V_{REV} est appliquée à ses bornes, équivalente à la tension de claquage de la diode plus la tension d'excès désirée pour être en mode Geiger. Lorsqu'un photon est absorbé dans la zone de déplétion et déclenche une avalanche, la tension au nœud V_A augmente. Lorsqu'elle atteint un niveau déterminé, le comparateur change d'état et active le circuit de contrôle logique, qui active à son tour le transistor d'éteuffement par le signal Quench et impose ainsi la tension V_{QUENCH} au nœud V_A . La tension V_{QUENCH} étant plus élevée que la tension d'excès appliquée à la photodiode, la tension aux bornes de celle-ci diminue sous la tension de claquage, éteuffant par le fait même le photodétecteur. Le transistor d'éteuffement reste activé jusqu'à ce que le temps mort établi par le circuit de contrôle logique soit écoulé, à la suite de quoi le transistor est désactivé, alors que le transistor de recharge est activé. La tension au nœud V_A est remise à la masse, repolarisant par la même occasion la photodiode au-delà de sa tension de claquage.

Comparativement au circuit d'éteuffement passif, le fait de venir imposer une tension à l'anode de la photodiode permet une fin d'éteuffement plus rapide. En contrepartie, le circuit actif nécessite un laps de temps pour réagir au courant avalanche et venir imposer la tension d'éteuffement [Tisa *et al.*, 2008b]. En effet, il faut considérer le temps que la tension au nœud V_A augmente au-dessus du niveau de seuil du comparateur, que celui-ci signale l'événement à la logique de contrôle et que cette dernière, à son tour, change l'état du transistor d'éteuffement. Pendant ce court délai de réaction, la photodiode en état d'avalanche libère énormément de charges, ce qui a pour effet d'augmenter le nombre de charges prises dans des pièges profonds, et donc le bruit postimpulsionnel, sans compter les chances de déclencher une photodiode environnante. De plus, en imposant un changement

de tension radical à l'anode de la photodiode pour l'étouffement, la capacité interne de la PAMP suit ce changement de tension et se vide instantanément. Ce phénomène provoque par conséquent une impulsion de courant pouvant abîmer les traces fournissant l'alimentation et entraîner des sauts de tension sur l'alimentation. Cependant, la recharge de la photodiode se fait quant à elle très rapidement en opposition au circuit d'étouffement passif. Cette vitesse de recharge permet d'avoir les mêmes conditions d'utilisation d'un événement à l'autre, résultant en une intensité de courant d'avalanche constante, puisque le circuit se replace rapidement dans les conditions désirées [Zappa *et al.*, 2003]. Cela permet également d'obtenir une meilleure probabilité de détection des PAMP. Le désavantage lié à cette rapidité de recharge provient de la possibilité qu'un photon soit absorbé dans la zone de déplétion et redéclenche la photodiode pendant l'exécution de la recharge, alors que l'impédance de la branche est faible. En effet, la résistance entre l'anode de la PAMP et la masse étant très petite pour permettre une recharge rapide, un grand courant d'avalanche peut être libéré, et ce, tout le temps que la logique de contrôle tient le circuit dans l'état de recharge. Il y a donc un risque de brûler le circuit et la photodiode si l'impédance du transistor de recharge est trop faible. Un avantage du circuit d'étouffement actif par rapport au circuit d'étouffement passif est son contrôle du temps mort et l'ajustement de sa durée [Rochas *et al.*, 2004]. Le bruit postimpulsionnel est grandement diminué de cette façon, permettant une meilleure utilisation des détecteurs. Le grand nombre de composants nécessaire à la réalisation de ce circuit le rend toutefois plus complexe et prend plus d'espace à sa réalisation que le circuit d'étouffement passif. De surcroît, la tension d'excès utilisée pour l'obtention du mode Geiger est limitée par la tension d'étouffement V_{QUENCH} disponible dans l'architecture du circuit et par la tension supportée par les transistors.

2.2.3 Circuit d'étouffement mixte

Tel que mentionné précédemment, les circuits d'étouffement passif et actif possèdent chacun leurs caractéristiques propres. Plusieurs combinaisons de ces deux types de circuit peuvent être réalisées afin de mettre l'accent sur certains aspects plus spécifiques, basés sur l'application pour laquelle les PAMP sont utilisées. Certains concepteurs peuvent opter pour un étouffement passif et une recharge active, d'autres pour un étouffement actif et une recharge passive. D'autres encore vont choisir un étouffement mixte passif et actif, et il peut en être de même pour la recharge. La composition la plus populaire est l'étouffement mixte passif et actif, combinée à une recharge active (figure 2.9), puisqu'elle joint les avantages de chacune des structures.

Figure 2.9 Circuit d'étouffement mixte [Tisa *et al.*, 2008b].

Dans cette approche, dès le déclenchement de l'avalanche, les résistances R_L et R_S débutent l'étouffement de la photodiode suivant le même principe que pour le PQC. Une fois la tension aux bornes de la résistance R_S suffisante pour déclencher le comparateur, un signal est envoyé à l'étage de la logique de contrôle qui se charge d'activer le transistor PMOS d'étouffement. La tension V_{QUENCH} est alors imposée à l'anode de la PAMP, terminant activement l'étouffement comme l'AQC. Une fois le délai de temps mort désiré terminé, la logique de contrôle désactive le transistor d'étouffement et active le transistor NMOS de recharge, permettant une recharge active. Lorsque le délai de la recharge est terminé, le NMOS est désactivé et la photodiode est prête pour une nouvelle utilisation.

Le circuit d'étouffement mixte, par l'union des deux types de circuit, possède à la fois les avantages du circuit d'étouffement passif et ceux du circuit actif. Les composants passifs agissent dès le début de l'avalanche et la fin de l'étouffement étant faite activement, le temps de l'étouffement est diminué. Cette combinaison minimise le nombre de charges impliquées dans l'avalanche, réduisant par le fait même le bruit postimpulsionnel [Tisa *et al.*, 2007], la diaphonie optique [Tisa *et al.*, 2008b] et la puissance dissipée. Tout comme le circuit d'étouffement actif, la recharge est très rapide et le temps mort est contrôlable. Toutefois, l'impulsion de courant lors de l'étouffement actif est toujours présente, tout comme la probabilité d'endommager l'électronique si un photon déclenche la photodiode pendant la recharge active. Le mélange des deux types de circuits augmente le nombre de composants nécessaires au fonctionnement du circuit d'étouffement. Sa dimension en est donc grandement affectée, augmentant l'espace occupé, sans compter la complexité de sa conception qui est également accrue. Finalement, tel le circuit d'étouffement actif, la ten-

sion d'excès utilisée pour la photodiode est limitée à la tension d'éteuffement V_{QUENCH} et par la tension maximale supportée par les transistors branchés à l'anode de la photodiode.

2.2.4 Circuit d'éteuffement avec comparateur de courant

Une alternative au circuit d'éteuffement actif est d'utiliser un comparateur de courant à l'entrée du circuit pour détecter le courant d'avalanche (figure 2.10). Un miroir de courant, formé par M1 et M3, est utilisé pour contrôler le transistor MQ en série avec la PAMP, afin d'augmenter sa résistance et d'éteuffer ainsi passivement la photodiode [Mita et Palumbo, 2008]. Un signal est en même temps envoyé à la logique du circuit pour avertir de l'arrivée de l'événement. La logique contrôle la recharge via le transistor à gauche du circuit, comme pour le circuit d'éteuffement actif.

Un autre circuit de comparateur de courant peut également être utilisé (figure 2.11) permettant d'ajuster le courant de seuil nécessaire au comparateur pour la détection d'un événement. Les branches par lesquelles se font activement l'éteuffement et la recharge de la PAMP sont absentes du schéma (figure 2.11) et peuvent varier en fonction du concepteur. Elles se connectent au nœud où se retrouve I_{IN} .

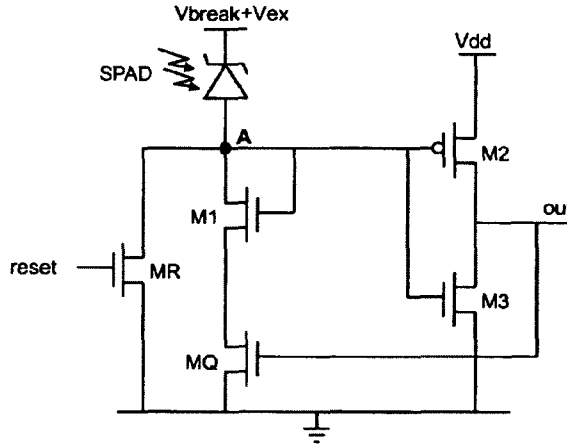


Figure 2.10 Circuit d'éteuffement avec comparateur de courant [Mita et Palumbo, 2008].

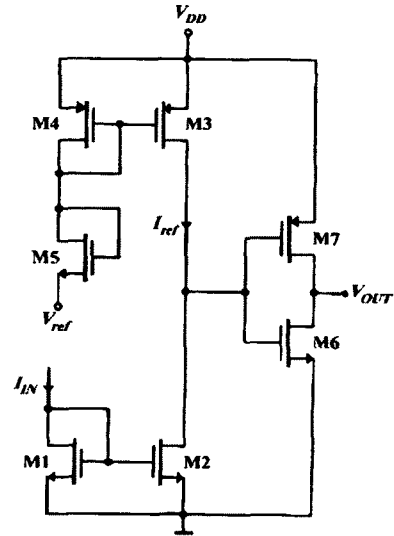


Figure 2.11 Circuit de comparateur de courant [Al-Ruwaihi et Noras, 1995].

Grâce au comparateur de courant, le courant d'avalanche est détecté plus rapidement qu'avec les comparateurs de tension. Par contre, la résistance offerte initialement à la

photodiode par les transistors est faible et augmente lentement pendant l'avalanche, permettant à un nombre important de charges de circuler à travers la zone de déplétion de la PAMP. Le temps d'éteuffement est plus long et le bruit postimpulsionnel est, par conséquent, plus élevé. Ce type de circuit vise plus particulièrement aux applications nécessitant une détection très rapide, mais peu fréquente. De cette façon, le circuit d'éteuffement peut garder la PAMP éteuffée plus longtemps pour permettre aux porteurs pris dans les pièges profonds de se libérer avant de recharger la photodiode.

2.2.5 Circuit d'éteuffement à charge variable

Les avantages du mélange des circuits d'éteuffement passif et actif sont très attrayants, mais l'espace requis à la réalisation d'un tel circuit est considérable. Pour pallier ce problème, le circuit d'éteuffement à charge variable (*Variable Load Quenching Circuit*, VLQC) a fait son apparition récemment [Tisa *et al.*, 2008b]. Sa structure, présentée à la figure 2.12, est accompagnée d'un schéma-bloc simplifié pour aider à sa compréhension (figure 2.13).

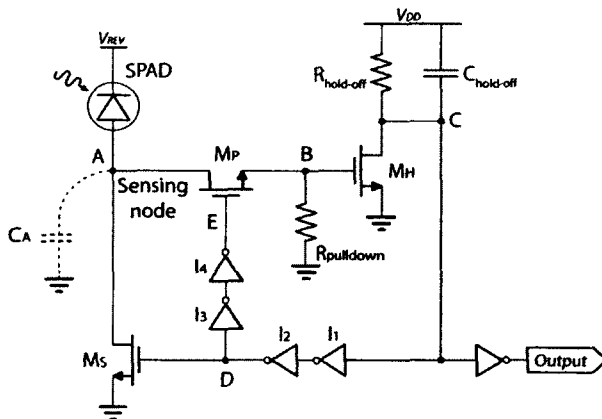


Figure 2.12 Circuit d'éteuffement à charge variable [Tisa *et al.*, 2008b].

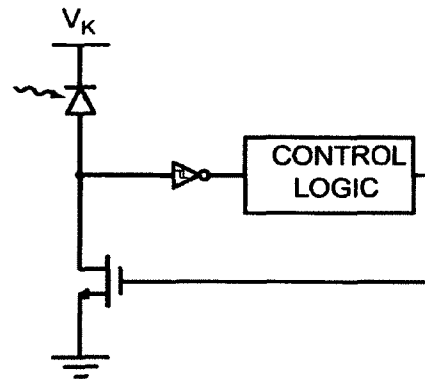


Figure 2.13 Schéma-bloc du circuit d'éteuffement à charge variable [Gallivanoni *et al.*, 2010].

À l'état initial, la tension d'alimentation V_{DD} est appliquée sur la grille du transistor NMOS M_S . Ce transistor étant actif, l'anode de la photodiode se retrouve à la masse. Lorsque la photodiode se déclenche, le courant circule via le transistor M_S . La tension au nœud A monte jusqu'à dépasser la tension de seuil du comparateur de tension. Ce rôle est joué par le transistor NMOS M_H et la tension de seuil de ce comparateur correspond à la tension de seuil V_{TH} du transistor. La résistance $R_{pull-down}$ est de valeur très élevée comparativement à l'impédance du transistor NMOS M_P . En raison du diviseur de tension résultant entre ces deux derniers, la tension au nœud B est sensiblement la même qu'au

nœud A. L'impédance totale de cette branche est également beaucoup plus grande que celle du transistor M_S , forçant ainsi le courant d'avalanche à circuler par cette dernière. Avec une tension fixe appliquée sur sa grille, le courant circulant par ce transistor est limité par le courant de saturation que le transistor peut laisser passer. Une fois le comparateur déclenché, les transistors M_S et M_P sont désactivés. Ce faisant, le transistor M_P déconnecte le comparateur de l'anode de la photodiode, le rendant donc insensible aux fluctuations sur le nœud A. Par la même occasion, cela assure que la tension sur la grille du transistor M_H ne soit pas trop élevée pour l'abîmer. Le transistor M_S étant lui aussi désactivé, son impédance devient dès lors énorme, finissant d'étouffer la photodiode encore plus rapidement. La logique de contrôle calcule ensuite le délai requis pour le temps mort, représenté à la figure 2.12 par la constante RC introduite par la résistante $R_{\text{hold-off}}$ et la capacité $C_{\text{hold-off}}$. Pendant ce temps mort, la tension au nœud B revient à la masse via la résistance R_{pulldown} . Au bout du temps mort, le transistor M_S est réactivé, permettant à la tension au nœud A de revenir à la masse. Les inverseurs I_1 et I_2 permettent un changement d'état rapide du transistor M_S lorsque la tension au nœud C est assez élevée. Après un délai introduit par les inverseurs I_3 et I_4 entre le nœud D et le nœud E, le transistor M_P est réactivé à son tour, afin de refaire la connexion entre l'anode de la PAMP et le transistor de comparaison M_H . Ce délai est nécessaire entre l'activation des deux transistors pour permettre à la tension au nœud A de descendre sous la tension de seuil du comparateur. En cas contraire, la logique de contrôle recommencerait la séquence d'étouffement, ainsi que le calcul du temps mort et la recharge ne pourrait jamais s'effectuer. Le circuit d'étouffement se retrouverait dans une boucle infinie et la photodiode ne serait par conséquent plus utilisable.

Le circuit d'étouffement à charge variable est donc une version particulière d'un circuit d'étouffement mixte, avec un étouffement passif et une recharge active. L'avantage de cet étouffement passif provient du fait que l'étouffement est fait par un transistor dont l'impédance peut varier. En effet, le changement d'état du transistor M_S permet une impédance très élevée à l'étouffement et plus faible en attente, ainsi qu'à la recharge, satisfaisant les besoins de chaque étape [Tisa *et al.*, 2008b]. Ce circuit reprend donc les avantages du circuit mixte passif-actif. L'étouffement passif permet une action d'étouffement dès le déclenchement de l'avalanche, tel que mentionné précédemment. Le changement d'impédance fournit une fin d'étouffement la plus rapide possible avec un étouffement passif, c'est-à-dire contraint à la constante RC interne de la photodiode [Cova *et al.*, 1996; Mita et Palumbo, 2008]. La recharge rapide réduit les probabilités d'absorber un photon pendant cet intervalle de temps. Cela permet également d'avoir une constance dans les conditions d'utilisation d'un événement à l'autre et de se trouver dans les conditions désirées pour

avoir un rendement optimal de la photodiode. Le contrôle du temps mort réduit le bruit postimpulsionnel et augmente par le fait même la probabilité de détection. En utilisant un transistor plutôt qu'une résistance pour faire l'étouffement, le circuit résultant est beaucoup plus compact que le circuit d'étouffement mixte. Cependant, le désavantage lié à la recharge active est encore une fois présent. En effet, si un photon est absorbé dans la zone de déplétion pendant la recharge et déclenche un courant d'avalanche, l'impédance du transistor M_S est très faible et un grand courant pourrait circuler et endommager l'électronique. Par contre, en faisant une bonne conception de ce transistor, il est possible de s'assurer que le courant ne soit pas trop dommageable dans un tel cas. De plus, si un photon déclenche la photodiode pendant la recharge, la tension au nœud A ne pourra revenir à la masse. Lorsque le transistor M_P sera réactivé, le comparateur de tension captera cette tension et recommencera immédiatement sa séquence d'étouffement, tout en envoyant un signal à l'électronique externe. La lecture de ces photons peut donc être faite, avec toutefois un retard entre le moment où le photon est absorbé et celui où le signal est envoyé à la sortie. Finalement, la tension d'excès pour polariser la photodiode est encore une fois limitée par la tension maximale pouvant être supportée par les drains des transistors M_S et M_P , mais n'est plus dépendante de la tension d'étouffement du type actif.

En résumé, le circuit d'étouffement passif est simple de conception et permet une réaction d'étouffement dès le départ de l'avalanche. Cependant, il doit faire un compromis entre sa vitesse d'étouffement et celle de recharge, en plus de ne pas contrôler le temps mort de la photodiode. Le circuit d'étouffement actif permet ce contrôle, mais requiert un temps de réaction du module de contrôle logique avant le début de l'étouffement. Le circuit d'étouffement avec comparateur de courant permet une détection rapide du courant d'avalanche, mais effectue un étouffement lent de la PAMP et s'applique plus particulièrement aux applications nécessitant des détections occasionnelles. Le circuit d'étouffement mixte combine les avantages des circuits passifs et actifs, mais requiert une grande surface pour sa conception. Le VLQC permet de réduire l'espace utilisé par le circuit, rendant ce circuit le plus attrayant de la littérature.

2.3 Transistor à drain étendu

Dépendamment du type de photodétecteurs utilisé, l'électronique qui s'y rattache peut devoir supporter une plus haute tension que ne le permet le procédé technologique utilisé. La PAMP en est un exemple, puisqu'en augmentant la tension d'excès à laquelle elle est soumise, sa probabilité de détection de photons augmente également. C'est donc un atout pour un circuit d'étouffement de pouvoir supporter à son entrée une tension supérieure

à la valeur maximale possible pour son électronique de lecture. Certaines techniques ont été développées afin de permettre aux transistors d'accepter une plus grande tension sur leur drain, comme les transistors à drain étendu. Plusieurs architectures, avec quelques variantes, ont été soumises dans la communauté scientifique, apportant tout à la fois divers acronymes pour les désigner : LDMOS (*Long Drain MOS* ou *Lateral Double diffused MOS*) [Aarts *et al.*, 2005; Wu *et al.*, 2006b], VDMOS (*Vertical Diffused MOS*) [Santos *et al.*, 2002; Tudor *et al.*, 2008], LDDMOS (*Low Doped Drain MOS* ou *Lightly Doped Drain MOS*) [Santos *et al.*, 2001; Shrivastava *et al.*, 2010], EDMOS (*Extended Drain MOS*) [Lee et Kwon, 1999; Sonsky et Heringa, 2007] ou à l'inverse DEMOS (*Drain-Extended MOS*) [Imoto *et al.*, 2009; LaFonteese *et al.*, 2009]. La nomenclature la plus retrouvée dans la littérature, soit DEMOS, sera utilisée ici. Un exemple de transistor de type NMOS et un de type PMOS d'architecture DEMOS sont présentés à la figure 2.14.

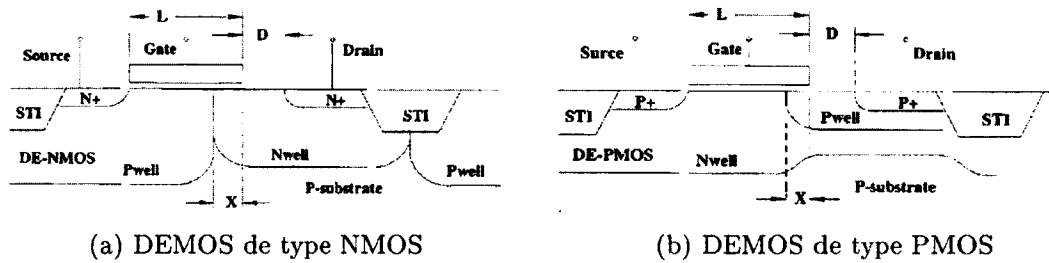


Figure 2.14 Exemple de transistors DEMOS [Bianchi *et al.*, 2008].

Normalement, la région dopée servant pour le drain d'un transistor est adjacente à la grille. La technique utilisée pour le DEMOS est d'éloigner le drain de la grille et d'utiliser une région moins dopée, mais de même type que le dopant de ce drain. Tel que démontré à la figure 2.14, un puits n, dopé de n^- (NWell), est utilisé comme extension du drain pour le NMOS, alors qu'un puits p, dopé de p^- (PWell), est utilisé pour le PMOS. Ce changement de dopage et la distance permettent de réduire le champ électrique à travers les lignes de champs équipotentielles (figure 2.15) créées avant d'atteindre la grille et la source du transistor [Mitros *et al.*, 2001]. Le drain du transistor peut donc supporter une plus grande tension sans occasionner de claquage avec la grille ou la source, tout en étant fabriqué avec le même procédé technologique. Selon certains tests menés dans la communauté scientifique, pour une technologie CMOS 0,18 μm , lorsque le transistor est opéré à 5 V, environ 70 % de la tension de drain est perdue dans l'extension. Toujours selon ces mêmes tests, la tension de claquage dépasse les 12 V pour cette architecture [Mitros *et al.*, 2001].

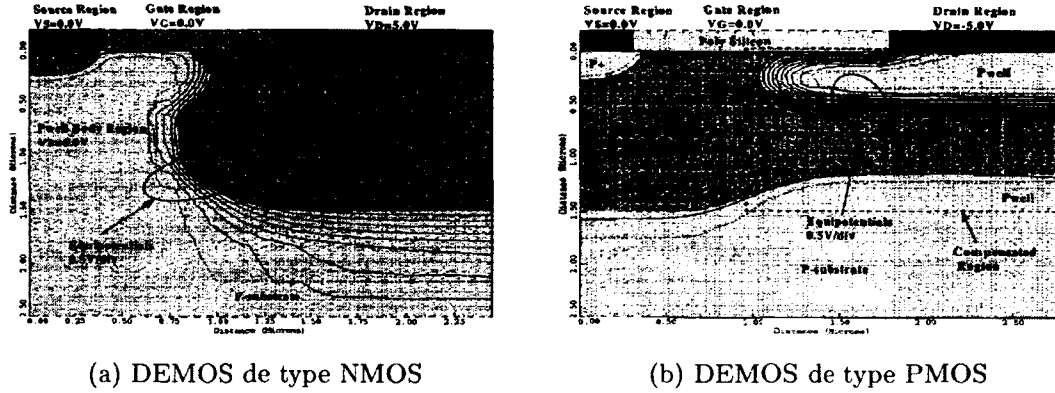


Figure 2.15 Lignes de champs équipotentielles de DEMOS [Mitros *et al.*, 2001].

Pour la conception d'une de ces architectures de transistor, trois dimensions, ayant chacune leur résistance, sont primordiales pour le rendement du transistor. Il y a la région du canal, la région d'accumulation et la région de dérive, représentées respectivement par R_{channel} , R_{ac} et R_{drift} (figure 2.16). La région de dérive est la zone étirée entre le drain et la grille sous l'oxyde de champ. La région d'accumulation se situe sous la grille, entre la zone dopée p^- et le début de l'oxyde de champ et la troisième région se retrouve dans la zone dopée p^- sous la grille entre la source et la forme allongée du drain. C'est à cet endroit que se forme le canal du transistor.

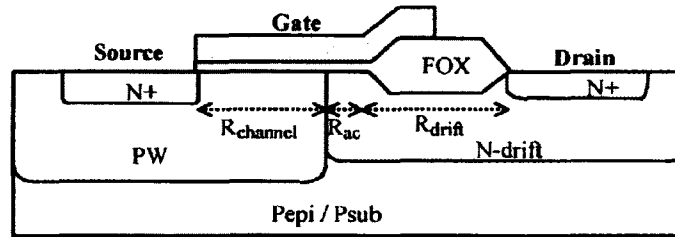


Figure 2.16 Les régions de canal, d'accumulation et de dérive dans un DEMOS [Wu *et al.*, 2006a].

En 2007, une nouvelle structure de DEMOS utilisant l'acronyme LDMOS a été proposée [Santos *et al.*, 2007]. Cette architecture novatrice allonge le puits n^- au-delà de la source de façon à englober la totalité du transistor, tout en entourant la source d'un puits p^- , lui-même à l'intérieur de ce même puits n^- (figure 2.17). La région de dérive sous l'oxyde de champ entre le drain et la grille, ainsi que la région d'accumulation sous la grille, mais toujours dans le puits n^- , sont toujours présentes telles que vues précédemment à la figure 2.16. La région du canal se situe, quant à elle, encore entre la région d'accumulation et la zone dopée n^+ de la source, dans le puits p^- . Le canal du transistor est donc créé

à l'intérieur de ce puits p^- , plutôt que directement dans le substrat de type p . Le puits p^- est polarisé par la région dopée p^+ à côté de la source, représentée par la lettre B sur la figure 2.17, acronyme du mot anglais *bulk*. L'auteur utilise cette terminologie afin de différencier le puits utilisé comme un substrat pour une architecture normale de NMOS et le véritable substrat. Dans le but de faciliter cette différenciation, le terme *puits-substrat* sera utilisé pour la suite de ce texte. Ce puits-substrat étant isolé du véritable substrat, cela permet d'assurer une certaine protection à l'électronique de faible tension près du DEMOS. En effet, si la tension appliquée au drain du transistor devient trop élevée pour le transistor et qu'un claquage survient, la jonction avec le puits-substrat flanchera avant celle avec le vrai substrat, empêchant donc les perturbations dans ce dernier. Un autre très grand avantage de cette structure est son efficacité en tant que transistor de passage. Étant donné que le puits-substrat est isolé du reste de l'électronique, il n'est pas obligatoire que sa tension soit à la masse. Par conséquent, lorsqu'une haute tension se situe au drain du DEMOS, il est possible d'élever la tension du puits-substrat, de la source et de la grille pour empêcher un claquage. Par exemple, un transistor de passage fabriqué dans un procédé 3,3 V pourrait avoir une tension à son drain de 20 V, une tension à sa source et à son puits-substrat de 10 V et avoir une tension de grille variant entre 10 V et 13,3 V. Ainsi, la tension de la grille à la source ou de la grille au puits-substrat varie seulement entre 0 V et 3,3 V, n'excédant pas, de cette façon, les conditions d'utilisation du procédé et empêchant un claquage de la grille à la source ou au puits-substrat. Un second DEMOS s'occuperait de la transition de 10 V à la masse. Grâce à cette nouvelle architecture de DEMOS, une tension de claquage plus élevée peut être atteinte que pour le DEMOS standard. Cette tension de claquage peut également être augmentée en allongeant la région de canal et en diminuant la région d'accumulation [Santos *et al.*, 2007]. Certains procédés de fabrication proposent des transistors à haute tension basés sur le principe de DEMOS. Cependant, ils requièrent des masques et des étapes de fabrication supplémentaires, imposant par le fait même des coûts additionnels. La dernière structure proposée n'utilisant qu'un puits n^- et un puits p^- , la fabrication ne nécessite aucun masque de plus ni aucun coût supplémentaire.

2.4 Intégration 3D

Les besoins grandissants pour les nouvelles applications ont poussé la communauté scientifique à chercher de nouvelles avenues au niveau de l'intégration de l'électronique. L'approche la plus en vogue est l'intégration 3D. Différentes techniques ont été développées dans le collage des gaufres pour obtenir ce type d'intégration, qui ont été divisées en quatre catégories : le collage adhésif et anodique, le collage direct de gaufres, le collage

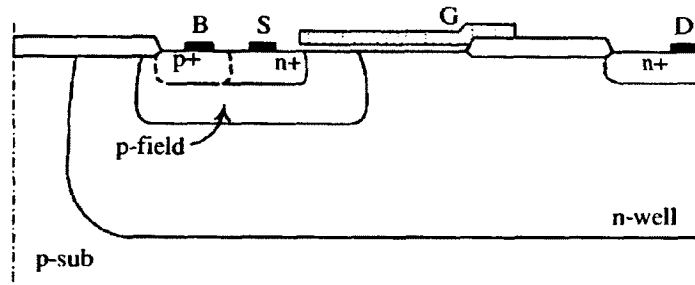


Figure 2.17 Nouvelle structure DEMOS avec le puits p^- isolé [Santos *et al.*, 2007].

métallique et le collage hybride métal/diélectrique [Ramm *et al.*, 2012]. Le procédé de fabrication Tezzaron/Chartered CMOS 130 nm 3D utilise un collage de gaufres cuivre à cuivre par thermocompression, une technique se retrouvant dans la troisième catégorie, soit le collage métallique. La couche de circuits d'étouffement et celle de traitement numérique de signaux se retrouvent par conséquent collées face à face, en utilisant des interconnexions verticales pour communiquer avec le monde extérieur. Ce collage est fait de gaufre à gaufre, mais certains collages peuvent être faits de puce à gaufre ou encore de puce à puce. C'est un collage puce à puce qui sera utilisé pour connecter la couche de PAMP avec la couche de circuits d'étouffement. Cette section présente donc ce type d'intégration, en commençant avec l'historique qui a mené l'industrie vers l'intégration 3D, avant de présenter les interconnexions verticales, lesquels se retrouveront entre les PAMP et les circuits d'étouffement.

2.4.1 Historique

En 1965, Gordon E. Moore définit une nouvelle règle qui guidera l'industrie pour les décennies suivantes. Dans son article *Cramming More Components onto Integrated Circuits*, paru dans le magazine *Electronics*, ce dernier observait que le nombre de transistors par circuit intégré doublait d'année en année depuis l'apparition de ce composant [Moore, 1965]. Il fit alors la prédiction que cette tendance allait se maintenir, suivant la courbe présentée à la figure 2.18. Il dut cependant réviser cette prédiction en 1975 en spéculant que le nombre de transistors par circuit intégré doublerait plutôt approximativement tous les deux ans. Cette prévision, qui fut surnommée avec les années la loi de Moore, s'avéra juste et devint une ligne directrice pour l'industrie.

Afin de suivre les prédictions de Moore, les fonderies doivent trouver de nouveaux procédés de fabrication pour réduire la taille des transistors. Malheureusement, l'industrie arrive à

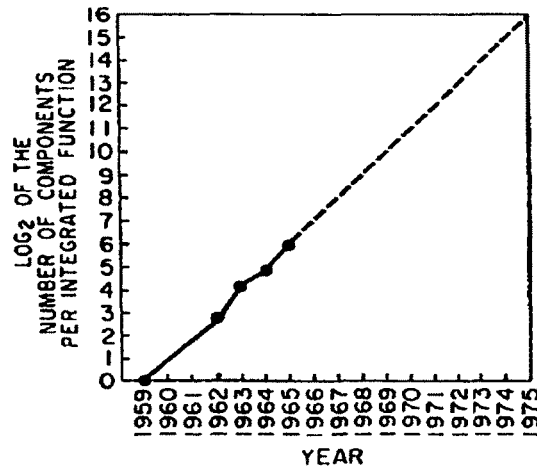


Figure 2.18 Loi de Moore [Moore, 1965].

un point où il devient de plus en plus difficile et coûteux de réduire la taille des composants. D'autre part, la miniaturisation des transistors perd un tant soit peu de son importance puisqu'il n'est plus le facteur dominant dans le temps de propagation des circuits intégrés. Tel que montré à la figure 2.19, pour les technologies plus petites que $0,35\ \mu\text{m}$, appelées DSM pour *Deep-Submicrometer*, les délais d'interconnexions deviennent le facteur le plus contraignant. La communauté scientifique doit par conséquent trouver une alternative à la loi de Moore, afin de réduire les délais dans les circuits intégrés tout en continuant d'augmenter leur densité.

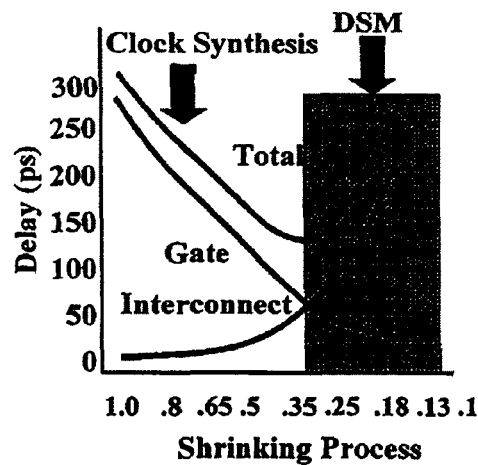


Figure 2.19 Délais en fonction du procédé de fabrication [Saleh *et al.*, 2000].
DSM : *Deep-Submicrometer*.

Un des moyens utilisés pour réduire ces délais est de réduire les distances d'interconnexion, diminuant par le fait même la capacité d'interconnexion et, ainsi, les délais engendrés. Cette diminution des distances d'interconnexion s'applique directement au bénéfice de l'utilisation de PAMP, où le circuit d'étouffement doit être placé le plus près possible de la photodiode. En réduisant cette distance, la capacité parasite entre eux est moindre, ce qui diminue le nombre de charges impliquées pour chaque événement ayant pour conséquence de réduire le bruit postimpulsionnel et la puissance consommée. De plus, cette plus courte distance et cette plus faible capacité parasite diminuent la gigue temporelle de la paire formée par la PAMP et le circuit d'étouffement. La première approche pour réduire les interconnexions fut de placer le circuit d'étouffement directement à côté de la photodiode, dans le même substrat [Finkelstein *et al.*, 2007; Niclass, 2008]. De cette façon, la capacité entre le circuit d'étouffement et la photodiode est réduite. Les délais d'interconnexions sont plus courts et les performances du circuit d'étouffement sont augmentées par rapport à la vitesse d'étouffement et au nombre de charges libérées pendant l'avalanche, ayant les répercussions discutées précédemment. Les figures 2.20 et 2.21 montrent une photodiode, un circuit d'étouffement à charge variable et de l'électronique de logique, ainsi que ce circuit en matrice. Cette architecture, présentée en 2009 par Simone Tisa, Fabrizio Guerrieri et Franco Zappa, réduit effectivement les délais d'interconnexion. Cependant, l'électronique utilise beaucoup d'espace, limitant grandement la surface sensible du détecteur. Pour ce pixel de $100 \times 100 \mu\text{m}^2$, seulement 3,14% de cette aire est photosensible. Cette solution ne s'avère donc pas être la plus optimale dans l'optique d'obtenir une grande probabilité de photodétection.

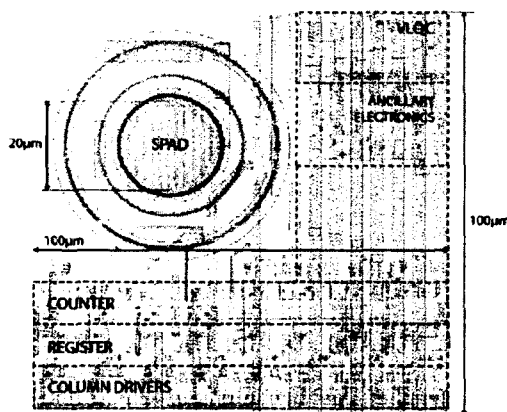


Figure 2.20 PAMP, circuit d'étouffement et électronique de logique [Tisa *et al.*, 2009].

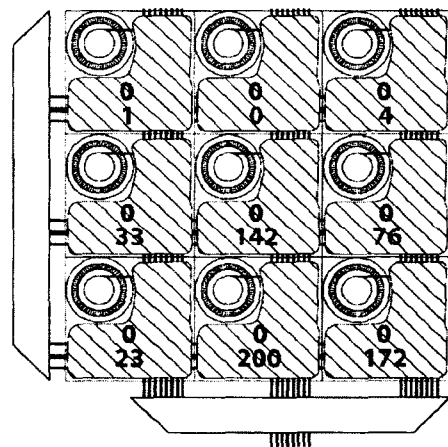


Figure 2.21 Matrice du circuit de la figure 2.20 [Guerrieri *et al.*, 2009].

2.4.2 Interconnexions verticales

Tel que mentionné précédemment, l'interconnexion entre la PAMP et son circuit d'étouffement doit être la plus courte possible, tout en minimisant l'impact sur le ratio de surface photosensible. L'intégration 3D est une des approches les plus en expansion et des plus enclines d'être la solution pour la prochaine génération de circuits intégrés. Le principe de fonctionnement est relativement simple : différents circuits intégrés sont empilés l'un sur l'autre et la connexion entre les différentes couches se fait par des interconnexions verticales (*Through Silicon Vias*, TSV). Il s'agit de cylindres, remplis de cuivre dans la plupart des cas, qui traversent le substrat permettant la connexion d'une couche à l'autre [Garrou *et al.*, 2008].

La réputée école du *Massachusetts Institute of Technology* (MIT) aux États-Unis a développé un système LADAR dans une architecture 3D, dont le montage est présenté à la figure 2.22 [Aull *et al.*, 2006]. La première couche contient une photodiode à avalanche opérée en mode Geiger. La seconde couche est composée du circuit de contrôle de la photodiode en technologie 3,3 V SOI et est connectée à la photodiode par le biais d'un TSV. La troisième couche contient l'électronique de traitement numérique de signaux, fabriquée en technologie 1,5 V SOI et est connectée à la seconde couche au moyen de cinq TSV. La connexion peut ainsi se faire entre les différentes couches, même si elles sont collées les unes aux autres verticalement. Cette solution permet d'obtenir une surface photosensible recouvrant la majeure partie de la première couche, comparativement à l'architecture 2D de la figure 2.20.



Figure 2.22 Système en intégration 3D [Aull *et al.*, 2006].

Plusieurs facteurs clés poussent l'industrie à mettre de l'avant la solution de l'intégration 3D, puisqu'ils répondent à différents critères nécessaires afin d'être acceptée dans le milieu des microtechnologies selon [Garrou *et al.*, 2008] :

- de meilleures performances électriques ;
- une réduction de la consommation de puissance et de bruit ;
- une amélioration du facteur de forme ;
- de faibles coûts ;
- plus de fonctionnalités.

Tout d'abord, les TSV permettent d'atteindre de meilleures performances électriques, puisque le raccourcissement de la longueur des interconnexions permet de réduire les délais causés par ces dernières. Pour les procédés de fabrication sous 0,35 μm , cette amélioration a un impact encore plus remarqué. De plus, pour la même aire de surface, l'intégration 3D permet une plus grande densité d'électronique, sans pour autant avoir recours à une diminution de taille des composants du circuit intégré, puisque l'empilement des couches permet un plus grand volume de transistors. Le raccourcissement des interconnexions réduit la capacité et la résistance de celles-ci, diminuant ainsi le bruit de couplage et la consommation de puissance. Pour le troisième critère, l'utilisation de l'intégration 3D permet d'intégrer directement la mémoire sous l'électronique de traitement numérique de signaux. La logique d'une puce électronique peut ainsi accéder directement à sa mémoire, augmentant de ce fait sa vitesse d'exécution. L'architecture 3D est également une solution peu coûteuse comparativement à chercher un nouveau procédé de fabrication des composants. Chaque fois qu'un nouveau procédé de fabrication est mis en place, les fonderies doivent modifier ou même acheter de nouvelles machines très coûteuses, alors que l'architecture 3D reprend les mêmes lignes de production. Finalement, le dernier critère est parfaitement comblé avec l'intégration 3D, puisqu'il permet les mélanges hétérogènes de technologies pour le même microsystème. Les couches sont fabriquées individuellement et assemblées ensemble par la suite, n'empêchant donc aucunement qu'elles soient de technologies différentes. Cela ouvre la porte à une multitude de possibilités de combinaisons qui s'avéraient impossibles avant l'apparition de l'architecture 3D.

Cependant, cette architecture comporte un désavantage majeur nécessitant un regard plus approfondi sur la question. En effet, les couches étant collées ensemble, la chaleur est confinée entre ceux-ci et sa dissipation en est par conséquent restreinte. Le concepteur doit donc porter une attention particulière, afin que la puissance dissipée soit minimale, réduisant

ainsi la quantité de chaleur à dissiper. Une température trop élevée pourrait endommager l'électronique, d'où le besoin pour l'industrie de rechercher de nouveaux moyens de refroidissement des couches. Une des approches en émergence utilise des microcanaux fluidiques pour résoudre cette problématique [Fourmigue *et al.*, 2013; Kim *et al.*, 2010].

2.5 Modèles de simulation

Outre la compréhension du fonctionnement des circuits d'étouffement et de leur relation avec les PAMP, de bons modèles représentant les photodiodes et les TSV sont nécessaires, afin de bien calibrer le circuit d'étouffement lors de sa conception. Grâce à ces modèles, il est plus facile d'estimer la gigue temporelle et la puissance consommée en fonction des paramètres choisis pour le circuit d'étouffement.

2.5.1 Photodiodes à avalanche

Avec les années, les circuits de modélisation de PAMP se sont améliorés et complexifiés. Les premiers modèles faisaient plutôt appel à la simplicité et ne nécessitaient que quelques composants. La figure 2.23 présente un modèle de base d'une PAMP, dans un contexte de circuit d'étouffement passif où R_S et R_L jouent leurs rôles présentés en section 2.2.1 [Cova *et al.*, 1996]. La capacité C_d représente la capacité de jonction de la photodiode, la résistance R_d exprime sa résistance interne de la PAMP et la source de tension symbolise sa tension de claquage. La capacité C_S équivaut à la capacité parasite de la cathode au substrat. Le commutateur est présent pour activer la branche de la résistance interne lors de l'arrivée d'un photon et pour la désactiver lorsque la tension aux bornes de la photodiode descend sous la tension de claquage. Le fonctionnement du modèle a été détaillé précédemment à la page 14 et ne sera donc pas répété ici. Dans le modèle suivant (figure 2.24), la capacité parasite a été remplacée par deux capacités parasites distinctes, soit une à la cathode et la seconde à l'anode. Cet ajout est important, puisqu'il influence directement la réponse entre la photodiode et le circuit d'étouffement, dépendamment du type de photodiode, soit n^+p ou p^+n , et de la borne de la PAMP à laquelle se fait l'étouffement, tel qu'expliqué à la section 2.2.

Le désavantage de ces modèles est le contrôle qui doit être fait sur le commutateur dans le premier cas et sur le transistor NMOS dans le deuxième cas. Le concepteur doit désactiver la branche résistive lorsque la tension aux bornes de la photodiode diminue sous la tension de claquage, sans compter que ces modèles ne considèrent pas les phénomènes d'automatisme et d'autoétouffement. Ces phénomènes dépendent du courant circulant une fois que

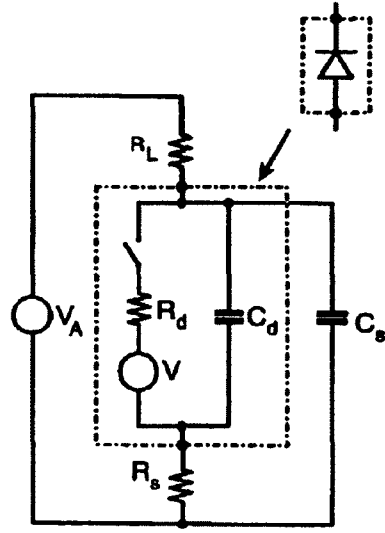


Figure 2.23 Modélisation simpliste d'une PAMP dans un PQC [Cova *et al.*, 1996].

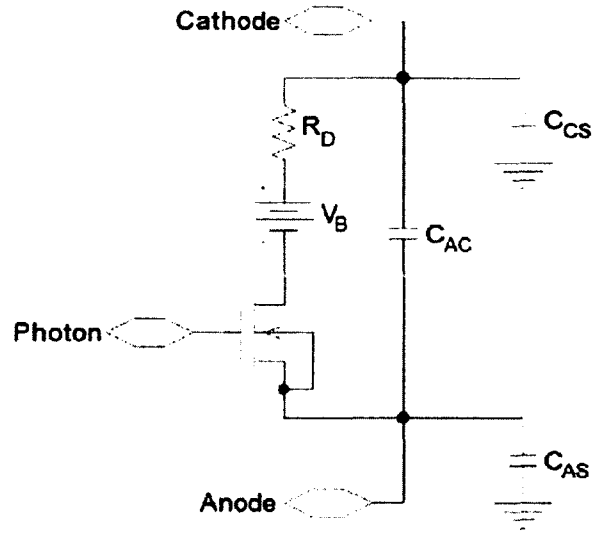


Figure 2.24 Modélisation simpliste d'une PAMP avec 2 capacités parasites [Dalla Mora *et al.*, 2007].

l'avalanche est déclenchée et qu'elle tend vers un état stable. Lorsque la photodiode se déclenche et que le commutateur se ferme, le courant immédiatement créé correspond au courant de la photodiode I_d , donné par l'équation :

$$I_d = \frac{V_{ex}}{R_d} \quad (2.1)$$

où V_{ex} équivaut à la tension d'excès aux bornes de la photodiode et R_d correspond à la résistance interne de la photodiode [Cova *et al.*, 1996]. Ce courant tend ensuite vers un état stable I_f , donné par l'équation :

$$I_f = \frac{V_{ex}}{R_d + R_L} \cong \frac{V_{ex}}{R_L} \quad (2.2)$$

R_L étant beaucoup plus grande que R_s , c'est la raison pour laquelle cette dernière a été négligée dans la première partie de l'équation 2.2. Le même raisonnement a été fait pour R_d entre la première et la seconde partie de cette même équation [Cova *et al.*, 1996]. Lorsque ce courant final I_f est plus grand qu'un courant de seuil d'environ 100 μA , valeur la plus répandue dans la littérature, l'avalanche s'automaintient [Dalla Mora *et al.*, 2007]. Dans le cas contraire, très peu de porteurs traversent la région d'avalanche, résultant en une forte probabilité qu'aucun porteur ne fasse d'ionisation par impact en traversant

le fort champ de la jonction. Cela mène par conséquent au second phénomène, celui de l'autoéteuffement de la photodiode. Afin de considérer ces importantes notions, le modèle de base de la PAMP a été modifié (figure 2.25).

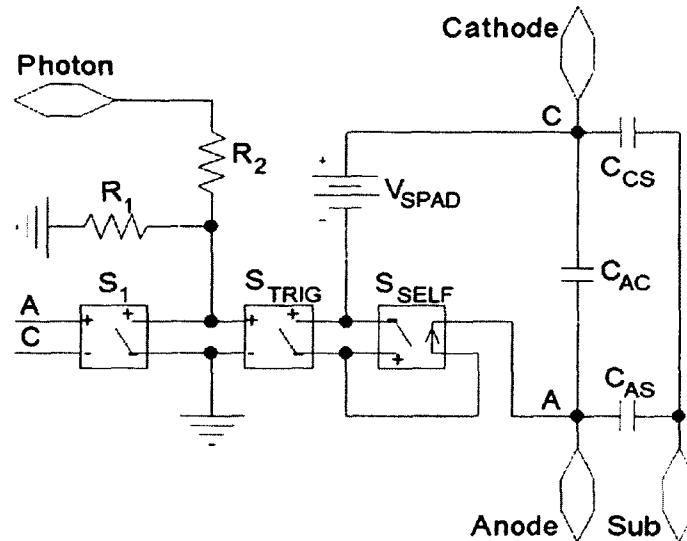


Figure 2.25 Modélisation d'une PAMP incluant la capacité à l'anode, ainsi que les phénomènes d'automaintien et d'autoéteuffement [Dalla Mora *et al.*, 2007].

En période d'attente, le commutateur de tension S_{TRIG} et le commutateur de courant S_{SELF} sont ouverts, permettant de développer la tension V_{AC} appliquée par le circuit externe entre la cathode et l'anode. Le commutateur de tension S_1 est tout d'abord fermé et ne s'ouvre que lorsque la tension V_{AC} dépasse la tension de claquage V_{SPAD} , empêchant ainsi la fermeture du commutateur S_{TRIG} et s'assurant que le déclenchement de la photodiode ne se fait que lorsqu'elle est en mode Geiger. Une fois la condition atteinte, le circuit est prêt à un déclenchement par l'entrée « Photon », les résistances R_1 et R_2 permettant d'utiliser autant une impulsion de tension que de courant comme stimuli. Une courte impulsion par cette entrée permet de fermer le commutateur S_{TRIG} et d'imiter le déclenchement de l'avalanche. Ce faisant, le courant engendré excède le seuil requis par le commutateur S_{SELF} , qui permet l'automaintien, et provoque sa fermeture en une fraction de seconde. Subséquemment, même si le signal fermant le commutateur S_{TRIG} n'est que de très courte durée, la fermeture de celui-ci est suffisamment longue pour démarrer le phénomène d'automaintien. La résistance interne de la photodiode R_d , en série avec la source représentant la tension de claquage des modèles précédents, est simulée avec V_{SPAD} grâce à un modèle *Orcad PSpice* du logiciel *Cadence* [Dalla Mora *et al.*, 2007]. Le même courant initial I_d est créé au déclenchement du modèle. Supposons qu'un circuit d'éteuffement

passif est branché aux bornes de ce modèle de PAMP. La tension augmentant aux bornes des résistances d'étouffement, celle entre la cathode et l'anode diminue, se rapprochant de la tension de claquage, représentée par la source V_{SPAD} . Selon l'équation 2.2, plus la tension d'excès diminue, plus le courant dans la photodiode diminue également, jusqu'à ce qu'elle atteigne le courant de seuil de 100 μA . L'autoétouffement de la photodiode est alors simulé à l'aide de l'ouverture du commutateur de courant S_{SELF} . La simulation de cet ajout au modèle est importante pour la conception du circuit d'étouffement, puisqu'une résistance d'étouffement trop faible pour un certain type de photodiode pourrait ne jamais provoquer l'autoétouffement et rester en avalanche grâce à l'automaintien. Cette modélisation convient pour la plupart des situations d'utilisation désirées, soit en polarisation inverse, en mode linéaire ou en mode Geiger. Cependant, le concepteur peut devoir simuler le circuit dans des conditions d'utilisation hors de cette plage. C'est pourquoi certains ajouts ont été faits au modèle précédent, résultant au modèle de la figure 2.26.

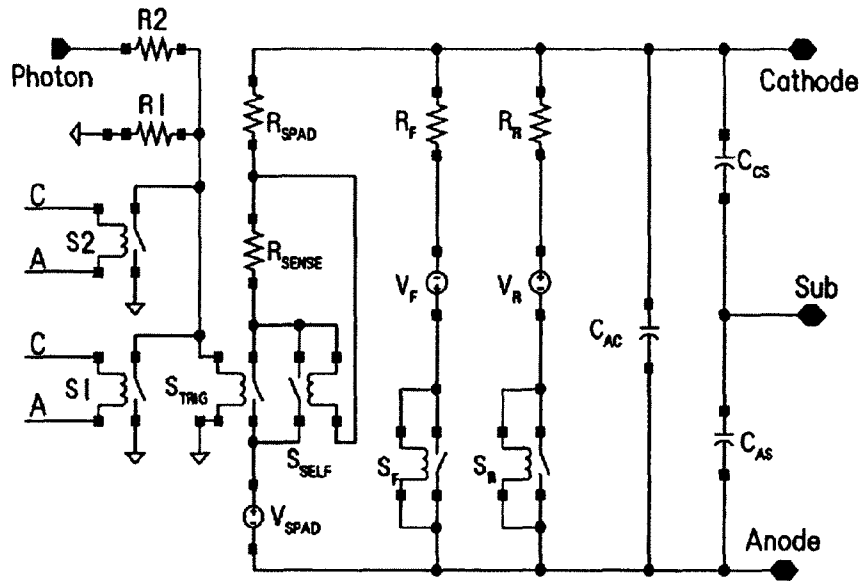


Figure 2.26 Modélisation d'une PAMP avec branche directe et second claquage [Zappa *et al.*, 2009].

Tout d'abord, une branche a été ajoutée pour les situations de polarisation en direct. Elle est constituée de trois composants, soit un commutateur de tension S_F , une source de tension V_F et une résistance R_F . La valeur de la source de tension V_F varie habituellement entre 0,6 V et 0,9 V, soit la tension nécessaire avant qu'une diode ne conduise lorsqu'elle est polarisée en direct [Zappa *et al.*, 2009]. Lorsque la tension à l'anode est plus élevée que celle de la cathode, mais plus faible que V_F , le commutateur S_F est ouvert et aucun courant ne circule. Une fois la tension V_F atteinte, le commutateur se ferme et permet au courant

de circuler, donnant une approximation du comportement d'une PAMP en polarisation directe. La seconde branche ajoutée au modèle comporte les mêmes éléments que pour la précédente, nommés S_R , V_R et R_R , en plus du commutateur S_2 . Cette addition au modèle permet d'ajouter une notion importante aux simulations : le second claquage. Lorsque la tension d'excès appliquée aux bornes de la photodiode est trop grande, le champ électrique devient trop fort pour la photodiode et un claquage se produit, permettant la circulation du courant. La photodiode est alors insensible aux photons puisqu'elle est d'elle-même en avalanche. La branche ajoutée au modèle fonctionne de la même façon que pour celle de la polarisation en direct, c'est-à-dire que le commutateur S_R est ouvert et qu'aucun courant ne circule par cette branche tant que la tension entre l'anode et la cathode n'excède pas celle du second claquage, symbolisée par la source de tension V_R . Une fois cette valeur dépassée, le commutateur est fermé permettant la circulation du courant, action imitée par le commutateur S_2 qui désactive toute sensibilité à la réception de nouveaux photons. Sur ce modèle, la résistance interne R_d , nommée R_{SPAD} à la figure 2.26, a été remplacée en série avec V_{SPAD} , en plus de R_{SENSE} servant à l'ajustement du seuil de l'automaintien et de l'autoétouffement.

Grâce à ces modèles, les simulations des circuits d'étouffement et de leur effet sur la photodiode à avalanche en mode Geiger se rapprochent grandement des résultats expérimentaux [Zappa *et al.*, 2009]. Une conception plus précise peut dès lors être faite, permettant d'obtenir des résultats plus près de ceux désirés.

2.5.2 Interconnexions verticales

Afin de considérer leurs impacts sur les circuits à concevoir, les TSV peuvent être modélisés par un circuit passif (figure 2.27). Le TSV étant fait de métal entouré d'oxyde servant d'isolant, sa capacité présente le même comportement que celui d'un transistor. En effet, une zone de déplétion s'y forme quand une tension est appliquée sur ce dernier et sa capacité varie entre une valeur maximale et une valeur minimale en fonction de cette tension. Sa plus haute valeur de capacité survient lorsqu'aucune tension n'est appliquée sur le transistor et équivaut à la capacité de l'isolant $C_{isolant}$, qui est l'oxyde dans le cas présent. Cette valeur est calculée grâce à l'équation 2.3. Sa plus faible capacité, donnée par l'équation 2.4, correspond à la combinaison de la capacité de l'isolant $C_{isolant}$ et de la capacité minimale de déplétion $C_{dep\ min}$ (équation 2.5), c'est-à-dire pour une zone de déplétion maximale [Katti *et al.*, 2010]. Dans ces équations, ϵ_{ox} correspond à la permittivité de l'oxyde, ϵ_{si} est la permittivité du silicium, l_{TSV} est la longueur du TSV et r_{metal} , r_{ox} , ainsi r_{max} sont différents rayons du TSV qui seront présentés à la section 3.2.2.

$$C_{\text{TSVmax}} = C_{\text{isolant}} = \frac{2\pi\epsilon_{\text{ox}}l_{\text{TSV}}}{\ln\left(\frac{r_{\text{ox}}}{r_{\text{metal}}}\right)} \quad (2.3)$$

$$C_{\text{TSVmin}} = \frac{C_{\text{isolant}}C_{\text{dep min}}}{C_{\text{isolant}} + C_{\text{dep min}}} \quad (2.4)$$

$$C_{\text{dep min}} = \frac{2\pi\epsilon_{\text{si}}l_{\text{TSV}}}{\ln\left(\frac{r_{\text{max}}}{r_{\text{ox}}}\right)} \quad (2.5)$$

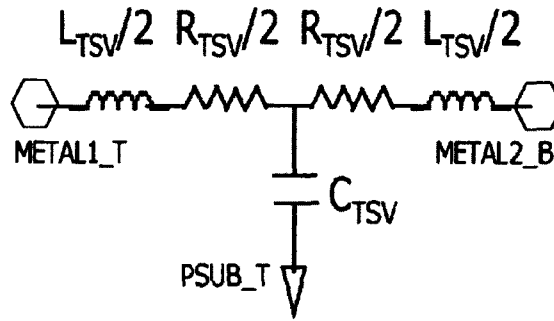


Figure 2.27 Modélisation d'un TSV [Katti *et al.*, 2010].

Après avoir déterminé les valeurs de capacité, il faut trouver celle de la résistance et celle de l'inductance. La valeur de la résistance du TSV est donnée par l'équation 2.6, où ρ est la résistivité du matériau, l_{TSV} est la longueur du TSV et r_{TSV} est son rayon [Katti *et al.*, 2010]. Finalement, l'inductance du TSV est calculée avec l'équation 2.7, où μ_0 représente la perméabilité dans le vide [Goldfarb et Pucel, 1991].

$$R_{\text{TSV}} = \frac{\rho l_{\text{TSV}}}{\pi r_{\text{TSV}}^2} \quad (2.6)$$

$$L_{\text{TSV}} = \frac{\mu_0}{2\pi} \left[l_{\text{TSV}} \ln \left(\frac{l_{\text{TSV}} + \sqrt{r_{\text{TSV}}^2 + l_{\text{TSV}}^2}}{r_{\text{TSV}}} \right) + \frac{3}{2} \left(r_{\text{TSV}} - \sqrt{r_{\text{TSV}}^2 + l_{\text{TSV}}^2} \right) \right] \quad (2.7)$$

En conclusion, une combinaison des PAMP et de l'intégration 3D semble la meilleure approche pour les modules de détection monophotonique. L'utilisation des PAMP comme photodétecteur permettra une meilleure sensibilité aux premiers photons et aidera à l'amélioration de la gigue temporelle. L'intégration 3D permettra de matricer ces détecteurs de petites tailles, sans électronique autour, augmentant ainsi le ratio de surface photosensible et améliorant la probabilité de détection de photons. L'intégration 3D permettra également une courte interconnexion entre la PAMP et le circuit d'étouffement, aidant à la

gigue temporelle. Le VLQC est le circuit d'étouffement qui fait le meilleur compromis par rapport au nombre de charges fournies par la PAMP, la rapidité d'étouffement et de recharge, le nombre de détections possibles, le contrôle du temps mort et l'espace utilisé. Par contre, le circuit est limité en tension par la technologie utilisée. Les transistors DE-MOS permettront de contourner cette limitation et ainsi améliorer le PDE. Finalement, les modèles de PAMP et de TSV permettront de bien calibrer les circuits d'étouffement grâce aux simulations avec le logiciel *Cadence*.

CHAPITRE 3

ANALYSE ET CONCEPTION

L'étude comparative des différents circuits d'étouffement existant dans la littérature ayant été faite, il est nécessaire par la suite de simuler les circuits les plus prometteurs. Le logiciel *Cadence* permet de faire ces simulations. Cependant, la simulation des circuits d'étouffement requiert la modélisation de PAMP et de TSV afin de prendre en considération les caractéristiques mutuelles de chacun.

Pour bien saisir les interactions entre les différents circuits du projet et les parties devant être simulées, la figure 3.1 montre l'intégration 3D du projet global développé. La première couche contient les photodiodes qui captent les photons. Étant de type p^+n , l'anode est, par conséquent, la terminaison la moins capacitive, tel qu'expliqué à la section 2.1.3. C'est la raison pour laquelle l'anode est branchée au TSV se rendant vers le circuit d'étouffement. La couche de photodiodes et celle des circuits d'étouffement sont collées dos à dos, nécessitant donc deux TSV, un dans chaque couche, pour se rendre d'un étage à l'autre. Le collage entre la deuxième et la troisième couche se fait face à face, par l'intermédiaire de plots d'interconnexion sur la sixième couche de métal. Finalement, la troisième couche nécessite des TSV pour faire le lien entre son électronique et la carte PCB, puisqu'elle y est collée de dos. Il faut également prendre en considération que cette puce servira pour la TEP. Par conséquent, il y aura un bloc de scintillateurs au-dessus de la couche de photodiodes. Chaque scintillateur a les dimensions de $1,2 \times 1,2 \text{ mm}^2$, soit $1,1 \times 1,1 \text{ mm}^2$ de structure cristalline, avec un isolant de $50 \text{ }\mu\text{m}$ en périphérie. Les photodiodes étant espacés de $50 \text{ }\mu\text{m}$ centre-à-centre, la matrice compte donc 484 PAMP, soit 22 par 22, afin de couvrir l'espace sous la structure cristalline du scintillateur. C'est pourquoi les circuits d'étouffement sous les photodiodes devront posséder autant d'éléments pour le même espace.

Suite aux simulations des circuits les plus prometteurs trouvés dans la littérature, certaines adaptations ont été faites afin d'offrir la combinaison s'adaptant le mieux au projet général. En premier lieu, le fonctionnement général de l'architecture choisie sera présenté. En second lieu, chacun des blocs de l'architecture sera examiné plus en détail afin de faire la lumière sur les choix de conception.

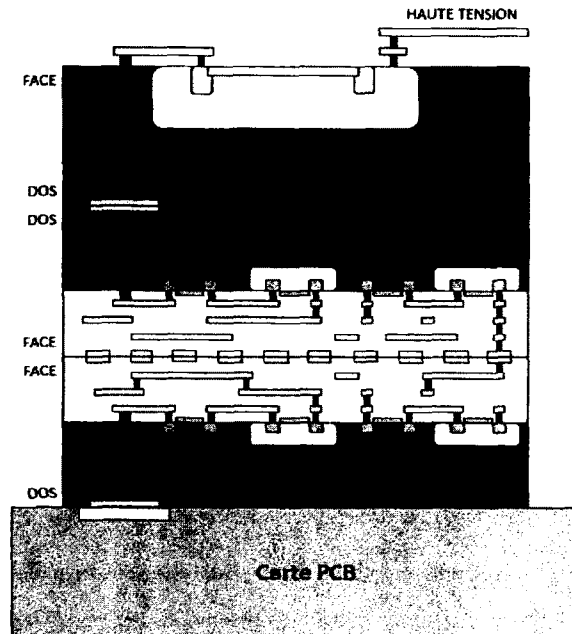


Figure 3.1 Intégration 3D du projet développé. Les 3 couches conçues sont intégrées ensemble verticalement avec la couche de PAMP au-dessus, la couche d'étouffement en second et la couche de traitement numérique de signaux en dessous.

3.1 Fonctionnement du circuit d'étouffement

La figure 3.2 présente un schéma-bloc du circuit afin de bien comprendre son fonctionnement. Ce schéma contient la plupart des modules nécessaires pour la simulation du circuit. Pour la couche 1, il faut modéliser la PAMP, ainsi que le TSV. Dans la même optique, il faut modéliser le TSV de la couche 2 sous l'électronique du circuit d'étouffement. Le circuit d'étouffement choisi est de type mixte, ayant un étouffement passif à charge variable et une recharge active, basé sur le VLQC vu à la section 2.2.5. Ce dernier contient l'électronique frontale, qui inclut une branche de détection et d'étouffement, une branche de recharge, ainsi qu'un comparateur de tension. Le circuit d'étouffement contient également une bascule, deux monostables, soit un pour le temps mort et l'autre pour le temps de recharge, la logique de contrôle, un générateur d'impulsions et des traducteurs de niveau qui ajustent la tension des bits de configuration. Un autre module indispensable au bon fonctionnement du circuit est absent du schéma, soit le module de l'amorce et des références de courant qui dirigent les monostables.

Afin de faciliter la compréhension du fonctionnement du circuit, les branches à l'entrée du circuit sont modélisées par des commutateurs. Lorsque le circuit est en attente d'un événe-

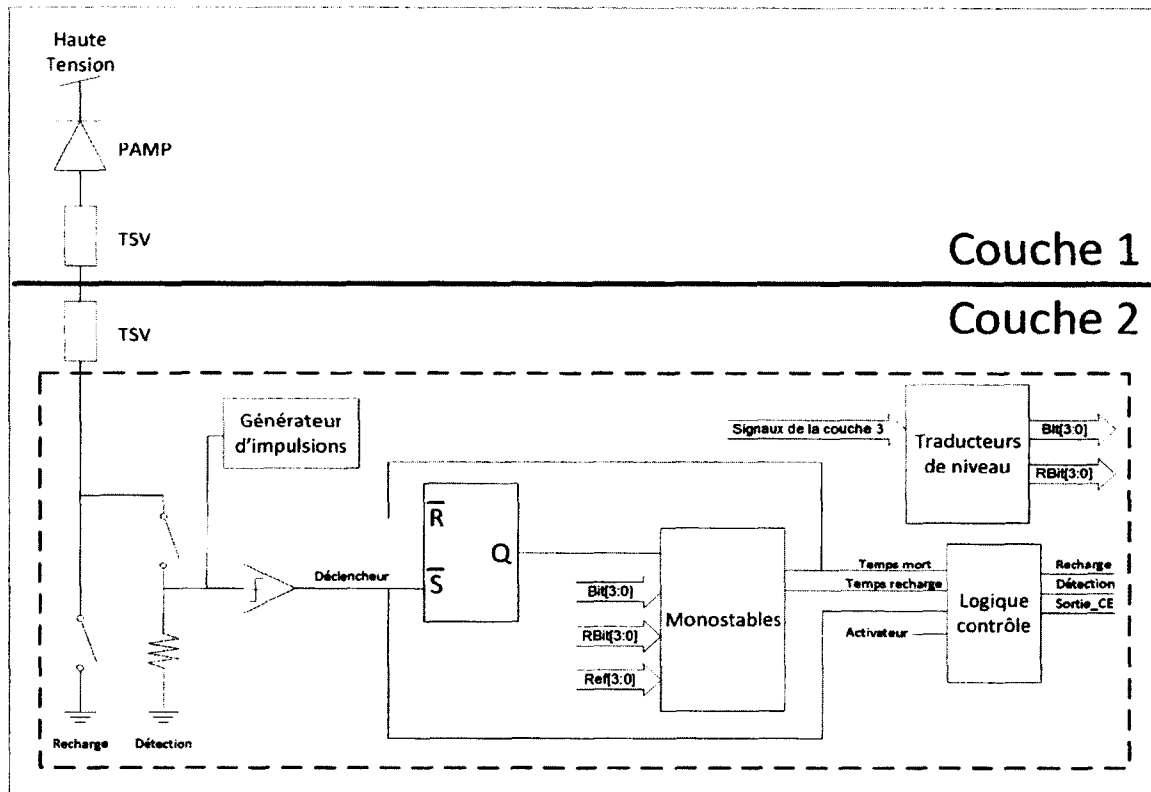


Figure 3.2 Schéma-bloc du circuit d'étouffement conçu. Les commutateurs à l'entrée représentent le circuit en mode d'étouffement.

ment, le commutateur de la branche de recharge est ouvert alors que celui de la branche de détection/étouffement est fermé. Cela assure que toutes les charges passeront uniquement par cette dernière. Lorsque la photodiode se déclenche, une avalanche de charges circule à travers le TSV de la première couche, fait de même avec le TSV de la seconde couche et arrive à l'entrée du circuit d'étouffement. Ces charges passent par la branche de détection et la tension monte au nœud du comparateur de tension, tout comme à l'anode de la photodiode. Une fois le seuil atteint, la bascule change d'état, ce qui a pour effet d'ouvrir le commutateur de la branche de détection/étouffement via la logique de contrôle, tout en démarrant le monostable du temps mort par la même occasion. De plus, la logique de contrôle envoie un signal de sortie à l'étage de traitement numérique de signaux, afin de l'avertir de l'événement. Les commutateurs des deux branches étant ouverts, représentant en fait deux énormes résistances, la tension à l'anode de la photodiode augmente rapidement jusqu'à ce que la différence de potentiel entre les bornes de la photodiode soit plus faible que sa tension de claquage, auquel cas l'avalanche est étouffée. La PAMP est gardée dans cet état pour une durée prédéterminée par le monostable utilisé pour générer

un temps mort. Lorsque le temps mort arrive à sa fin, la bascule revient à son état initial grâce à la logique de contrôle, ferme le commutateur de la branche de recharge, permettant à la tension à l'anode de revenir à la masse et, ainsi, de retrouver le potentiel de sa tension de claquage et de la tension d'excès appliquée à ses bornes. Au même moment, le monostable générant le temps de recharge débute son compte à rebours. Le temps de recharge arrivé à terme, le commutateur de la branche de recharge est de nouveau ouvert, alors que celui de la branche de détection/étouffement revient en position fermée. Celle-ci n'est pas fermée pendant la recharge, sans quoi la tension à l'entrée du comparateur serait plus haute que la tension de seuil et le circuit d'étouffement se redéclencherait aussitôt. Le changement d'état des commutateurs fait, le circuit est prêt à recevoir l'événement suivant. Le générateur d'impulsions permet d'imiter le déclenchement d'une PAMP et de s'assurer de la fonctionnalité du circuit, ainsi que sa communication avec la couche de traitement numérique de signaux avant de débiter l'acquisition de données. En effet, ce dernier déclenche le comparateur de tension à l'entrée du circuit d'étouffement, alors que la logique numérique vérifie que sa sortie change d'état pour chaque impulsion soumise par le générateur. Toutefois, en se connectant à l'entrée du comparateur de tension et non directement à l'entrée du circuit, cette structure ne teste pas la fonctionnalité des deux transistors d'entrée, soit le transistor de détection/étouffement et celui de recharge. Le générateur d'impulsion a été placé à cet endroit afin de ne pas ajouter de capacité parasite à ce nœud critique et permet de s'assurer du bon fonctionnement de tout le restant du circuit, ainsi que de la bonne communication avec la couche de traitement numérique de signaux. La figure 3.3 présente tous les modules nécessaires pour la simulation du circuit d'étouffement. La figure 3.4, tirée de simulations avec le logiciel *Cadence*, complète l'information donnée en affichant l'allure des courbes pour chacun des signaux principaux à l'intérieur du circuit. Ce graphique pourra aider à bien comprendre le fonctionnement de chacun des modules lorsqu'ils seront expliqués plus en détail dans les sections à venir.

3.2 Simulations

3.2.1 Modélisation de la PAMP

Afin de modéliser une PAMP, le modèle de la figure 2.26 est privilégié puisqu'il s'agit du plus complet trouvé dans la littérature. Les informations nécessaires à la simulation, telles la capacité de jonction, la capacité anode-substrat, ainsi que la capacité cathode-substrat, ont été fournies par les fabricants et ne peuvent pas être divulguées. Elles sont toutefois présentées dans l'annexe confidentielle B.

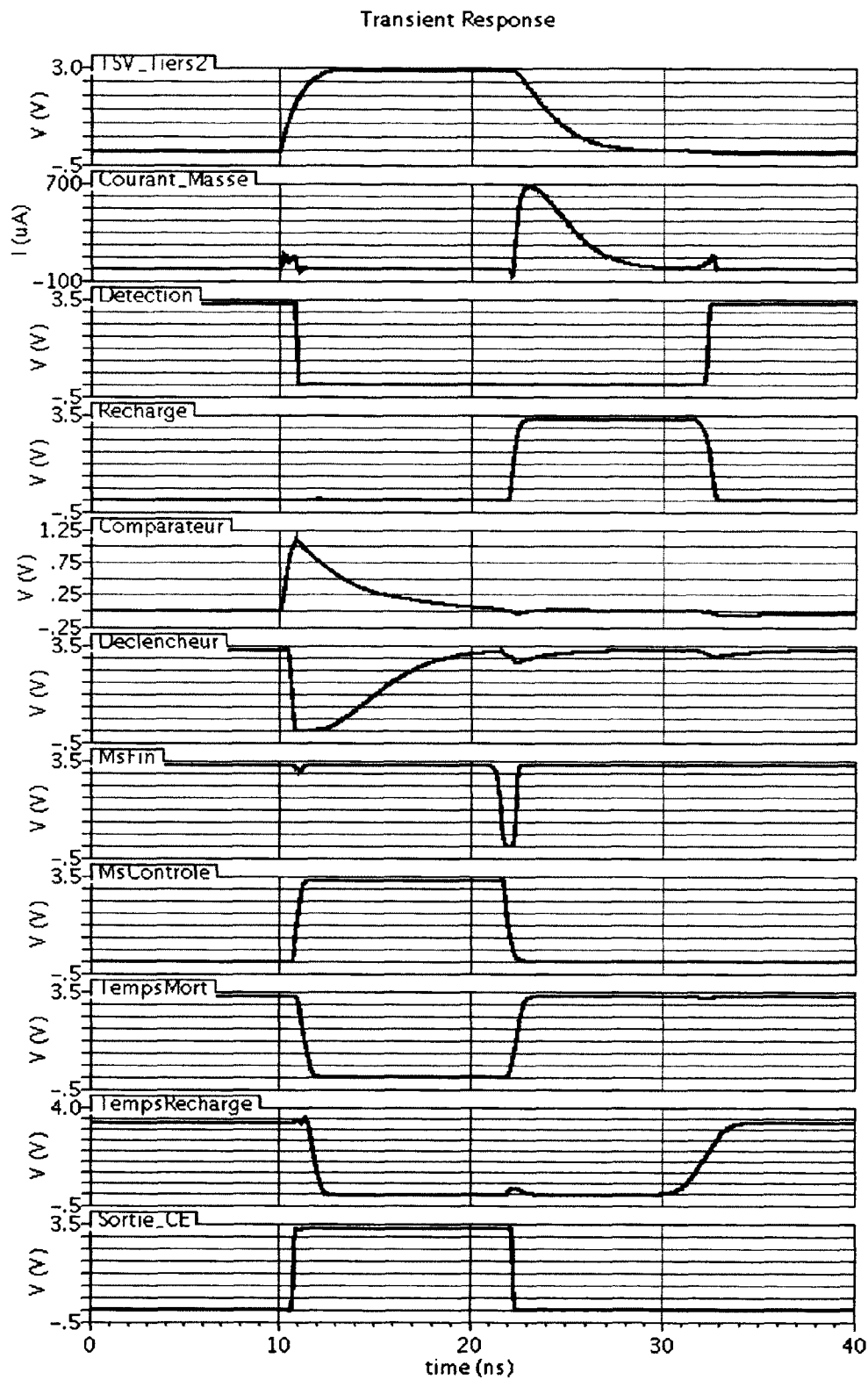


Figure 3.4 Courbes des principaux signaux simulés avec le logiciel *Cadence*. Pour ces simulations, la PAMP a été déclenchée à 10 ns.

3.2.2 Modélisation du TSV

Afin de considérer l'effet des TSV, les simulations utilisent le modèle présenté à la section 2.5.2. Les équations 2.3 et 2.4 présentées en page 2.3 expriment respectivement la capacité maximale et la capacité minimale du TSV. Cette dernière est représentée en fonction de la capacité de l'oxyde C_{isolant} et de la capacité minimale de déplétion $C_{\text{dep min}}$, elle-même calculée avec l'équation 2.5. Ces équations utilisent trois rayons différents faisant partie du TSV. Ces rayons se retrouvent à la figure 3.5, où r_{metal} (a) est le rayon de la partie métallique du TSV, r_{ox} (b) ajoute la couche d'oxyde servant d'isolant, alors que r_{max} (c) considère en plus la zone de déplétion maximale pouvant être créée autour du TSV [Katti *et al.*, 2010]. La capacité du TSV changeant en fonction de la largeur de la zone de déplétion, les capacités limites maximales et minimales sont calculées afin de simuler le circuit dans les cas extrêmes. Rappelons que deux types de TSV se retrouvent entre la photodiode à avalanche et le circuit d'étouffement, soit un reposant à l'intérieur de chacune des couches. Les TSV de la couche de photodiodes étant développés post-traitement par le Centre de Recherche en Nanofabrication et Nanocaractérisation (CRN2) de l'Université de Sherbrooke et ceux de la couche d'étouffement développés par Tezzaron, il faut calculer les caractéristiques de chacun des deux types de TSV séparément. Cependant, encore ici, certaines informations, telles les tailles et la concentration des dopants, ne peuvent être divulguées ici, mais se retrouvent dans l'annexe confidentielle B.

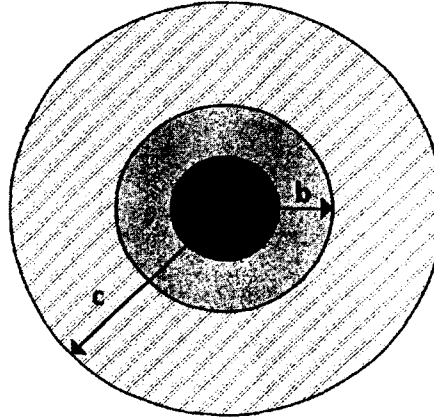


Figure 3.5 Rayons d'un TSV : (a) r_{metal} ; (b) r_{ox} ; (c) r_{max} .

En débutant avec l'équation 2.3, la variable ϵ_{ox} correspond à la permittivité de l'oxyde, soit sa constante diélectrique relative ϵ_r multipliée par la permittivité du vide ϵ_0 . Cela équivaut à $3,9 \times \epsilon_0$, c'est-à-dire $3,9 \times 8,85 \times 10^{-12}$ F/m [Katti *et al.*, 2010; Streetman et Banerjee, 2005]. La longueur et le diamètre des TSV de chacune des couches ayant été fournis par les fabricants, il ne reste que r_{metal} à déterminer. En effet, la longueur du TSV

correspond à l_{TSV} et le diamètre donné divisé par deux résulte en r_{ox} , puisque la taille du TSV donné par le fabricant considère le métal et l'isolant autour. La valeur de r_{metal} se trouve en enlevant l'épaisseur d'isolant utilisé autour du TSV du rayon r_{ox} . Le fabricant Tezzaron utilisant une épaisseur d'oxyde de 1000 Å [Garrou *et al.*, 2008], cette valeur a été utilisée pour le calcul des TSV des deux couches, puisque cette information n'était pas disponible pour la couche des photodiodes en date du calcul. La capacité maximale du TSV est ainsi trouvée.

Pour l'équation 2.5, un peu plus de démarches sont nécessaires afin d'arriver au résultat. Dans les variables nécessaires à la résolution de cette équation, la longueur du TSV et r_{ox} ont toutes deux été présentées au calcul précédent, alors que la permittivité du silicium ϵ_{si} équivaut à $11,8 \times \epsilon_0$, soit $11,8 \times 8,85 \times 10^{-14}$ F/cm [Streetman et Banerjee, 2005]. Cependant, pour avoir r_{max} , il faut tout d'abord connaître la largeur maximale de la zone de déplétion (W). Cette dernière se calcule par les équations suivantes [Streetman et Banerjee, 2005] :

$$\Phi_s = \frac{2kT \ln \left(\frac{N_a}{n_i} \right)}{q} \quad (3.1)$$

$$W = \sqrt{\frac{2\epsilon_{\text{si}}\Phi_s}{qN_a}} \quad (3.2)$$

où l'expression kT provient de la constante de Boltzmann ($8,62 \times 10^{-5}$ eV/K) à la température de la pièce, soit $2,59 \times 10^{-2}$ eV. La concentration intrinsèque des électrons pour le silicium à la température de la pièce n_i vaut $1,5 \times 10^{10}$ cm $^{-3}$, alors que la charge électronique q équivaut à $1,6 \times 10^{-19}$ C. La concentration du dopant N_a a été fournie par la compagnie Dalsa qui fabrique la couche des photodiodes. Cette donnée étant inconnue pour la couche de Tezzaron, la même valeur de dopant a été approximée. Connaissant la valeur de la largeur de la zone de déplétion maximale, cela permet de trouver r_{max} en ajoutant cette valeur au rayon avec oxyde calculé précédemment. Toutes les valeurs sont dès lors connues pour résoudre l'équation 2.5 et, par la suite, l'équation 2.4, permettant ainsi de trouver les valeurs de $C_{\text{dep min}}$ et de la capacité minimale du TSV C_{TSV} .

La résistance équivalente du TSV se calcule avec l'équation 2.6. Pour un TSV fait de cuivre, la résistivité est de 16,8 nΩ×m et le rayon utilisé correspond au r_{metal} du calcul précédent [Katti *et al.*, 2010].

Finalement, l'inductance équivalente du TSV se calcule par l'équation 2.7, où la perméabilité dans le vide est donnée par $4\pi \times 10^{-7}$ H/m [Goldfarb et Pucel, 1991; Katti *et al.*, 2010]. Encore ici, le rayon utilisé correspond à r_{metal} calculé précédemment.

3.2.3 Électronique frontale 3,3V

Telle qu'étudiée au chapitre 2, l'architecture du circuit d'éteignement à charge variable fait le meilleur compromis entre les différents aspects du circuit d'éteignement et revient au plus avantageux. Toutefois, en ayant la même tension appliquée sur sa grille, le transistor à l'entrée du circuit se retrouve dans le même état pour l'éteignement et pour la recharge. Cela implique donc un compromis à faire entre une faible impédance, permettant une recharge rapide, mais un début d'éteignement lent, et une haute impédance, permettant un début d'éteignement rapide, mais une recharge lente. L'utilisation de deux branches, soit une pour l'éteignement et l'autre pour la recharge, permet d'éviter de faire ce compromis et sera l'approche utilisée (figure 3.6). Le désavantage de cette approche réside toutefois en la capacité parasite qui augmente à l'anode de la photodiode, un prix tout de même peu cher payé pour avoir un éteignement et une recharge rapides. En effet, même si la capacité parasite au nœud est un peu plus élevée, l'éteignement plus rapide de la photodiode permet de diminuer le nombre total de charges circulant pendant la période d'éteignement. Il en est de même pour la recharge ; malgré le fait que la capacité à recharger soit plus grande, le transistor de recharge étant conçu spécifiquement pour cette tâche, le temps de recharge est plus court et la photodiode est prête pour un nouvel événement plus tôt. Afin de ne pas avoir trop de transistors branchés à l'anode de la PAMP, le comparateur de tension est placé entre le transistor d'éteignement MN1 et un second en cascode, doit MN2. De cette façon, l'anode ne voit que deux drains de transistors, tout comme pour le circuit original. Le transistor d'éteignement doit être petit pour offrir une faible capacité parasite à l'anode de la photodiode. Dans l'optique que le comparateur de tension déclenche le plus rapidement possible, le second transistor doit être d'une plus haute impédance que le transistor d'éteignement, un diviseur de tension résultant à ce nœud. Les simulations ont permis de déterminer la taille optimale du transistor MN2 en cascode. Les valeurs des tailles des transistors ne peuvent être mentionnées, mais se retrouvent dans l'annexe confidentielle A. En effet, l'impédance du transistor doit être assez grande, afin de limiter le nombre de charges lors de l'avalanche et permettre une augmentation rapide de la tension à son drain pour un déclenchement rapide du comparateur de tension. En revanche, son impédance ne doit pas être trop grande, puisqu'une fois le transistor d'éteignement éteint, la tension à l'entrée du comparateur de tension doit diminuer assez rapidement,

de façon à revenir sous le seuil du comparateur avant de recommencer à faire la détection d'événements.

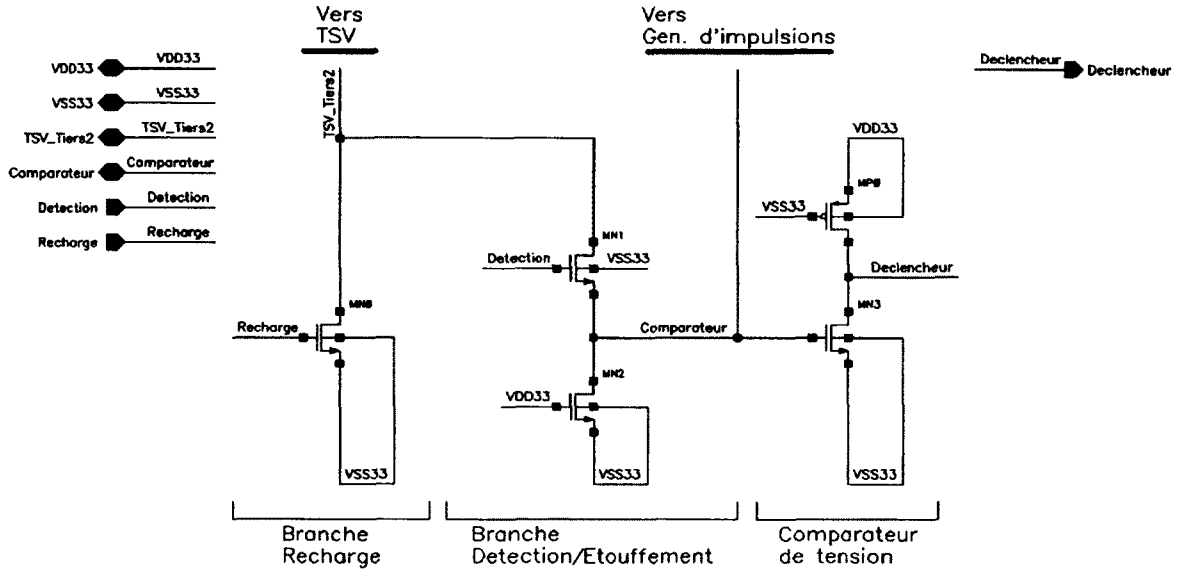


Figure 3.6 Schéma de l'électronique frontale formant l'entrée du circuit d'étouffement.

Le comparateur de tension est réalisé par seulement deux transistors. La tension de seuil du comparateur correspond à la tension de seuil V_{TH} du transistor NMOS MN3. Son impédance doit être faible afin d'amener la tension à l'entrée de la bascule à la masse et que la transition du signal se fasse rapidement. Le transistor PMOS MP0, de haute impédance, sert à remettre la tension à VDD à l'entrée de la bascule une fois le transistor MN3 éteint. Par contre, tout comme pour le transistor MN2 en cascode avec le transistor d'étouffement MN1, ce transistor PMOS ne doit pas être de trop grande impédance, afin de permettre un retour à la tension initial avant la détection suivante. Lors de la conception, il est primordial de s'assurer que l'électronique de logique de contrôle éteint le transistor de détection/étouffement assez rapidement pour ne pas qu'une trop haute tension ne se retrouve sur la grille du transistor MN3 du comparateur de tension, sans quoi il pourrait claquer.

La recharge de la photodiode se fait pour sa part par le biais d'un seul transistor NMOS, soit MN0. Certains éléments ont dû être pris en considération lors de sa conception, dont tout d'abord d'être de faible impédance pour permettre une recharge rapide. Il ne doit toutefois pas être de taille minimale pour tolérer une grande densité de charges. En effet, selon le modèle de simulation de PAMP utilisé, plus de charges circulent par l'alimentation du circuit d'étouffement à la recharge de la photodiode qu'à sa décharge. Tel qu'illustré à

la figure 3.7, la plupart des charges accumulées par la capacité anode-cathode C_{AC} servent à l'équilibrer à la décharge. Selon le modèle, l'alimentation fournit des charges au fur et à mesure que la tension s'élève à l'anode en raison du diviseur de tension formé par la résistance interne de la PAMP R_{SPAD} et la résistance d'étouffement R_{CE} , chargeant ainsi en partie la capacité parasite anode-substrat C_{AS} et la capacité des TSV C_{TSV} . La masse de l'alimentation du circuit d'étouffement ne voit passer que les charges provenant de la haute tension passant par sa branche de détection/étouffement, qui est représentée par la résistance R_{CE} , en fonction de la tension montante à l'anode de la photodiode. À la recharge par contre, le commutateur de la photodiode étant ouvert, toutes les charges accumulées par les trois capacités C_{AC} , C_{AS} et C_{TSV} se drainent par la branche de recharge du circuit d'étouffement au fur et à mesure que la tension à l'anode de la PAMP revient à la masse.

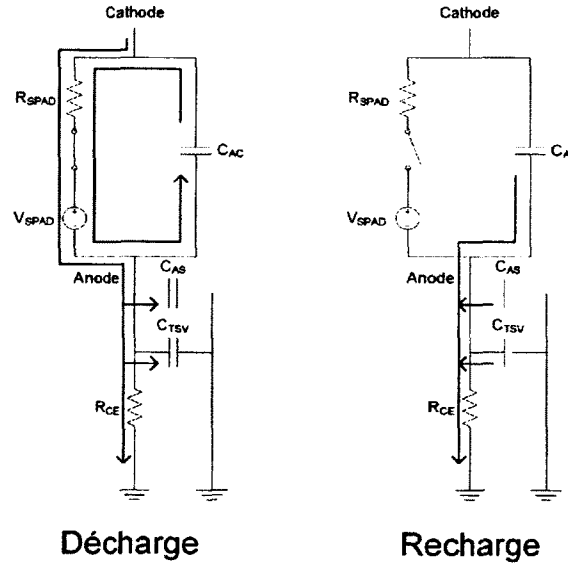


Figure 3.7 Déplacement des charges pour la décharge et la recharge.

De surcroît, deux raisons incitent à ce que son impédance ne soit pas trop faible. La première provient du fait que le circuit est matricé. Une recharge trop rapide crée une impulsion de courant qui doit être fournie par l'alimentation. Lors d'un événement TEP, la majorité des PAMP de la matrice de 22×22 circuits se déclenche au même moment. La somme de ces impulsions de courant provoque une baisse de tension dans la matrice due à la résistivité des traces. De plus, si le nombre demandé de charges est trop grand pour ce que les traces peuvent supporter étant donné leur largeur restreinte dans la matrice, cela pourrait endommager les traces et le circuit. La seconde raison provient de la possibilité d'un déclenchement de la photodiode à avalanche pendant la recharge. La branche de

détection étant désactivée pendant cette étape, et ce, pour la durée déterminée par le monostable, l'avalanche s'emballerait suivant l'hypothèse d'une impédance du transistor trop faible. Cela aurait comme conséquence, encore une fois, de détruire la puce. Une certaine impédance minimale doit donc être établie pour limiter ces cas. Toutefois, advenant le cas d'un déclenchement de la photodiode pendant la recharge, comme l'entrée du comparateur revient à la masse pendant ce temps, le circuit d'étouffement se redéclenche dès que la branche de détection est réactivée. Cela permet par conséquent à l'étage de traitement numérique de signaux de détecter cette absorption, malgré le retard qui l'affecte, ayant pour avantage d'améliorer la résolution en énergie.

3.2.4 Électronique frontale de haute tension

L'électronique frontale pour la haute tension est identique au circuit de la section précédente, à l'exception du transistor de recharge, ainsi que celui de détection/étouffement. Cependant, les transistors de haute tension étant de conception personnalisée, aucun modèle ne permet de simuler le comportement exact de ce circuit. Il a donc fallu trouver une solution qui allait s'en rapprocher le plus près possible. Le procédé Chartered/Tezzaron contient une trousse de transistors de haute tension (5 V), qui n'était pas disponible pour la fabrication, mais dont les modèles de simulation étaient accessibles. Ces transistors sont également de type DEMOS, mais de conception différente. Ces modèles se rapprochant le plus des transistors à concevoir, ceux-ci ont été utilisés pour approximer les tailles des transistors et le comportement du circuit d'étouffement contenant ceux-ci. Les mêmes limitations de courant que pour le précédent circuit ont été prises en considération pour la conception de ces transistors.

3.2.5 Bascule

Une bascule a été positionnée à la sortie du comparateur de tension, afin de s'assurer que l'état du circuit reste le même pour toute la durée du temps mort et qu'il ne soit pas influencé par les changements d'état à l'entrée du circuit. Cette bascule a dû être conçue à la main, puisqu'aucune cellule propriétaire ARM n'était disponible avec les transistors 3,3 V dans le procédé de fabrication utilisé. La bascule est formée de deux portes logiques NON-ET (*NAND*), la première servant à changer l'état du signal en sortie et la seconde servant à ramener la sortie de la bascule à l'état initial (figure 3.8). Le signal *Declencheur* (figure 3.4) de la sortie du comparateur de tension vient donc activer la sortie de la bascule. Cette dernière n'est désactivée que lorsqu'elle reçoit le signal *MsFin* provenant du monostable du temps mort une fois le délai écoulé. Les transistors PMOS des portes

NON-ET de la bascule ont été conçus avec des largeurs W plus grandes, étant donné que le signal de sortie de la bascule est branché à huit grilles de transistors et doit par conséquent fournir plus de charges en raison de leur capacité.

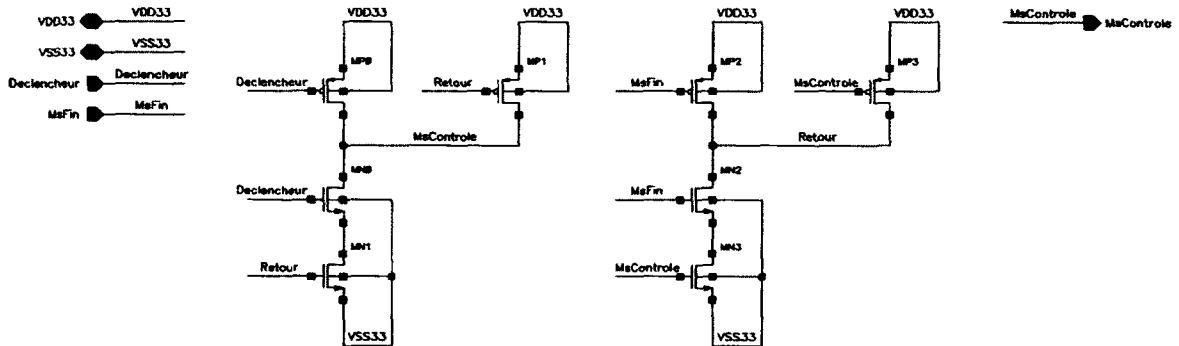


Figure 3.8 Schéma des deux portes NON-ET formant la bascule.

3.2.6 Monostable du temps mort

Ce monostable permet d'ajuster le temps mort pendant lequel la photodiode reste étouffée. Son architecture est montrée à la figure 3.9. À l'état initial, le nœud *MsTMInterieur* à l'entrée de la bascule de Schmitt (*Schmitt Trigger*) est polarisé à la tension d'alimentation. Ainsi, lorsque la sortie de la bascule *MsControle* vue à la section précédente passe du niveau bas au niveau haut, tel que montré à la figure 3.4, cela a pour effet de drainer les charges accumulées par les capacités du monostable. Lorsque la tension à ce nœud atteint la tension de seuil de 1,18 V de la bascule de Schmitt, celle-ci modifie son signal de sortie *MsFin*, qui doit être inversé avant de désactiver la bascule de l'étage précédent, signalant ainsi la fin du temps mort. Ces charges peuvent s'écouler via quatre branches différentes, dont trois d'entre elles sont contrôlées par des bits de configuration. Cela permet donc huit combinaisons différentes de ces branches, dont chacune possède sa propre référence de courant, lui permettant de cette façon de laisser passer différents courants d'une branche à l'autre. La vitesse à laquelle le nœud se décharge dépend donc de la combinaison de branches utilisée. De plus, un quatrième bit de contrôle est utilisé afin d'ajouter une plus grande capacité au nœud par l'intermédiaire d'un commutateur. En ajoutant des capacités à ce nœud, plus de charges doivent être drainées, ce qui a pour effet d'augmenter le temps nécessaire pour atteindre le seuil de référence pour les mêmes combinaisons de branches. Il en résulte donc seize combinaisons possibles grâce à ces quatre bits qui sont propres à chaque circuit, présentés au tableau 3.1.

Les transistors MN8 et MP0 branchés à l'entrée du monostable servent d'inverseur pour démarrer et arrêter le monostable. Le PMOS est de taille plus grande, afin de permettre

Tableau 3.1 Les 16 possibilités de temps morts simulés.

| Bit3 | Bit2 | Bit1 | Bit0 | Temps (ns) |
|------|------|------|------|------------|
| 0 | 0 | 0 | 0 | 849,95 |
| 0 | 0 | 0 | 1 | 502,51 |
| 0 | 0 | 1 | 0 | 407,04 |
| 0 | 0 | 1 | 1 | 317,02 |
| 0 | 1 | 0 | 0 | 115,91 |
| 0 | 1 | 0 | 1 | 110,84 |
| 0 | 1 | 1 | 0 | 107,46 |
| 0 | 1 | 1 | 1 | 103,40 |
| 1 | 0 | 0 | 0 | 73,32 |
| 1 | 0 | 0 | 1 | 45,25 |
| 1 | 0 | 1 | 0 | 37,20 |
| 1 | 0 | 1 | 1 | 29,95 |
| 1 | 1 | 0 | 0 | 11,70 |
| 1 | 1 | 0 | 1 | 11,50 |
| 1 | 1 | 1 | 0 | 11,30 |
| 1 | 1 | 1 | 1 | 11,20 |

une recharge rapide des capacités une fois le temps mort terminé. La vitesse de recharge de ces capacités est cruciale, puisqu'elles doivent être complètement rechargées avant l'événement suivant, sans quoi le temps mort calculé par le monostable serait inconstant. Chacune des branches est connectée au nœud *MsTMInterieur* à l'entrée de la bascule de Schmitt par l'intermédiaire d'un transistor NMOS de taille minimale, soient MN0, MN1, MN2 et MN3. Trois d'entre eux, les transistors MN1, MN2 et MN3, sont activés par les bits de configuration avant même le début de l'acquisition des détections et gardent le même état pendant tout le temps de l'acquisition. La quatrième branche est contrôlée directement par l'alimentation du circuit et est par conséquent toujours active. Cette configuration a été choisie pour plusieurs raisons. Tout d'abord, comme le monostable doit avoir au minimum une branche active pour pouvoir fonctionner, cela permet d'avoir huit configurations possibles pour les trois bits utilisés. En effet, si cette quatrième branche était absente, la configuration où tous les bits sont au niveau bas serait impossible. Ensuite, cette branche étant toujours active, la présence de ce transistor n'est pas nécessaire pour son fonctionnement. Cependant, il a été ajouté afin que tous les transistors NMOS utilisés pour les références de courant, soient MN4, MN5, MN6 et MN7, se retrouvent dans les mêmes conditions, c'est-à-dire un transistor NMOS à leur drain et un second à leur source. Ces transistors de référence sont tous de taille minimale et la grille de chacun d'entre eux est référencée à son propre miroir de courant provenant du bloc de références de courant qui sera étudié à la prochaine section. Afin d'accumuler les charges au nœud

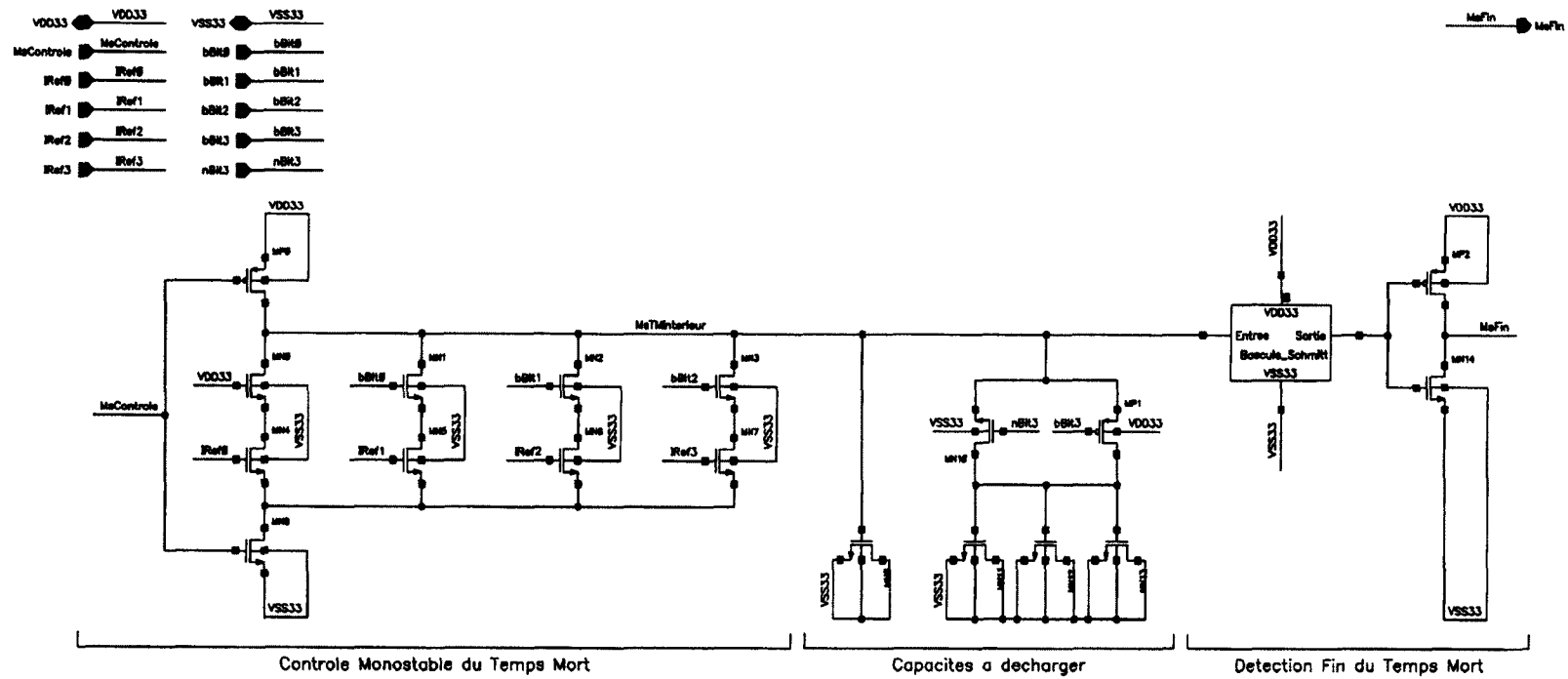


Figure 3.9 Schéma du monostable du temps mort.

d'intérêt, le circuit utilise les grilles des transistors NMOS MN9, MN11, MN12 et MN13, comme capacité MOSCAP considérant les contraintes en espace du circuit et que les capacités métal-isolant-métal (MIMCAP) sont beaucoup plus larges pour une même valeur de capacité. Leur différence de capacité par aire est présentée dans l'annexe confidentielle B.

La bascule de Schmitt (figure 3.10) a été choisie pour de multiples raisons. Pour commencer, les seuils de détection haut et bas sont ajustables selon la conception de la bascule, ce qui permet d'avoir une plus grande plage dynamique de tension avant le seuil et, par conséquent, une plus grande différence de temps pour une même valeur de courant et de capacité, selon l'équation $I = C \frac{\Delta V}{\Delta t}$. En effet, avec le seuil de détection à 1,18 V lors d'un passage de la tension d'alimentation vers la masse à l'entrée, une plage de 2,12 V est donc disponible avant la transition, comparativement à 1,65 V ($V_{DD}/2$) pour un inverseur régulier. Un autre avantage s'applique pour les longs temps morts, pour lesquels les charges sont drainées plus lentement, adoucissant la pente de décharge. En considérant l'effet du bruit sur la pente, il serait possible que près du seuil de détection, la sortie vacille en fonction du bruit. La bascule de Schmitt permet de s'immuniser de cet effet, puisqu'une fois le seuil de détection dépassé, le second seuil de détection se trouve à une tension trop éloignée pour que le bruit ne puissent changer la valeur de sortie.

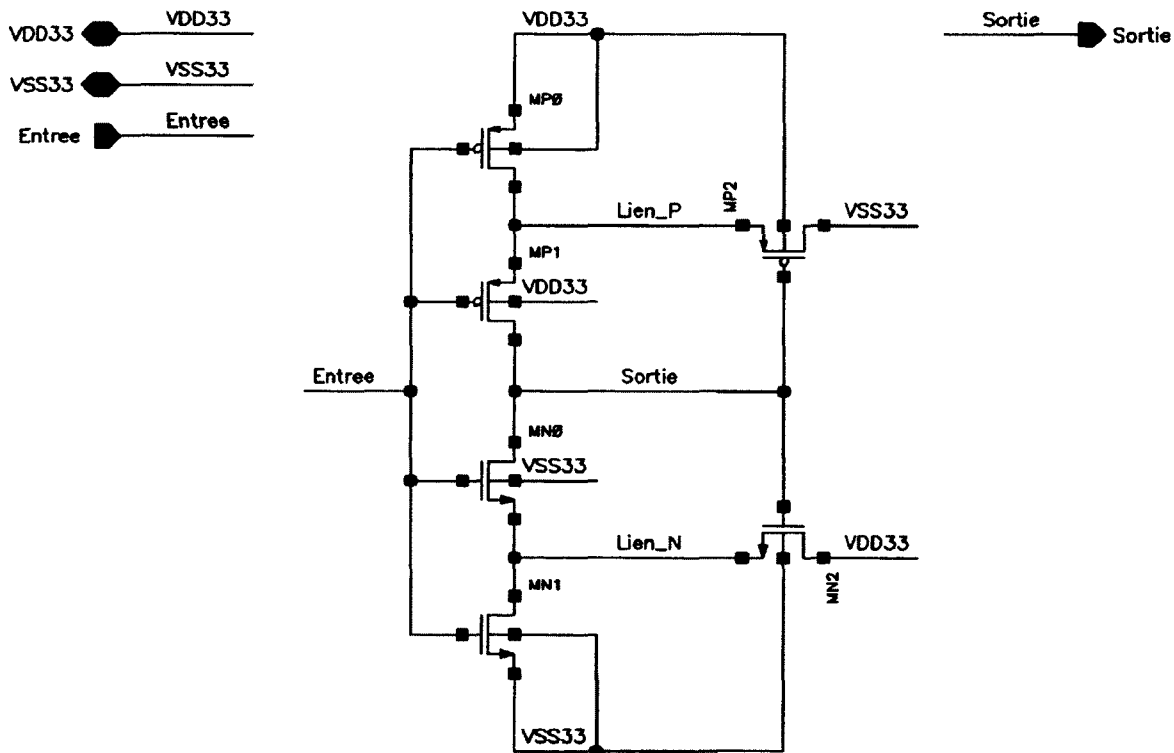


Figure 3.10 Schéma de la bascule de Schmitt.

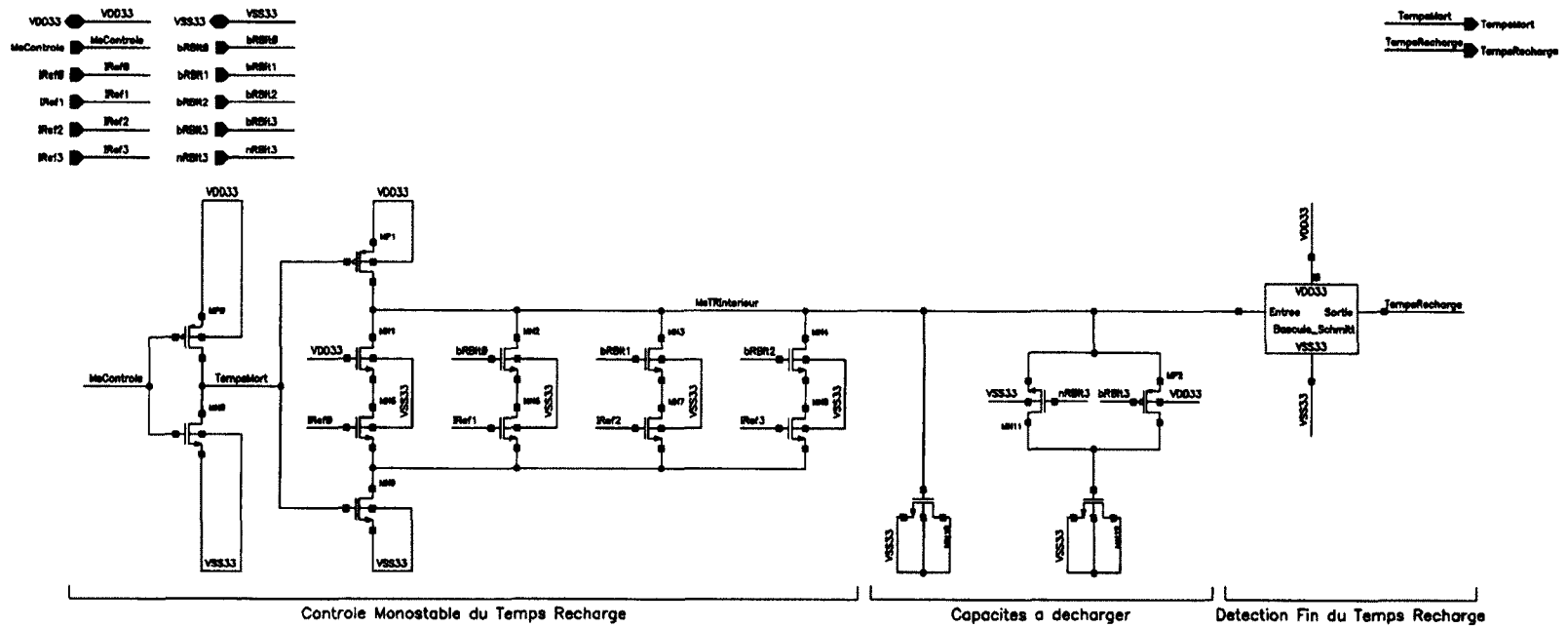
3.2.7 Monostable du temps de recharge

Le monostable présenté à cette section détermine la période de temps allouée pour la recharge de la photodiode. Ce circuit fonctionne de la même façon que le monostable présenté à la section précédente, à la différence qu'à l'état initial, les capacités sont déchargées. Grâce à l'inverseur à son entrée (figure 3.11), ce monostable utilise le temps mort pour charger ses capacités. La fin du temps mort donne le signal de départ pour ce monostable. Ainsi, le temps de recharge ne commence à se calculer que lorsque le temps mort est terminé, peu importe sa valeur. Les temps simulés pour chacune des combinaisons possibles sont présentés au tableau 3.2.

Tableau 3.2 Les 16 possibilités de temps de recharge simulés.

| RBit3 | RBit2 | RBit1 | RBit0 | Temps (ns) |
|-------|-------|-------|-------|------------|
| 0 | 0 | 0 | 0 | 108,48 |
| 0 | 0 | 0 | 1 | 66,52 |
| 0 | 0 | 1 | 0 | 54,54 |
| 0 | 0 | 1 | 1 | 43,70 |
| 0 | 1 | 0 | 0 | 17,26 |
| 0 | 1 | 0 | 1 | 16,81 |
| 0 | 1 | 1 | 0 | 16,35 |
| 0 | 1 | 1 | 1 | 16,01 |
| 1 | 0 | 0 | 0 | 62,54 |
| 1 | 0 | 0 | 1 | 39,30 |
| 1 | 0 | 1 | 0 | 32,40 |
| 1 | 0 | 1 | 1 | 26,47 |
| 1 | 1 | 0 | 0 | 10,75 |
| 1 | 1 | 0 | 1 | 10,60 |
| 1 | 1 | 1 | 0 | 10,32 |
| 1 | 1 | 1 | 1 | 10,24 |

L'électronique d'entrée est la même que pour le monostable précédent, à l'exception de l'inverseur mentionné plus tôt. Le monostable de recharge utilisant le temps mort pour charger ses capacités, il est impératif que le transistor PMOS utilisé permette une charge complète de ces dernières en une période de temps plus courte que le plus petit temps mort. Les transistors NMOS servant de références de courant sont couplés aux mêmes miroirs de courant que le monostable du temps mort. Le contrôle des délais se fait par l'intermédiaire de quatre bits, les trois premiers activant les branches de drainage, alors que le quatrième bit et son inverse activent ou désactivent l'accès à la capacité supplémentaire. Ces quatre bits spécifiques à la recharge s'appliquent à tous les monostables de recharge de la matrice de 22 circuits de long par 22 circuits de large, comparativement aux monostables de temps



mort qui sont contrôlés individuellement. Ce choix de conception découle de deux raisons. La première s'avère pour limiter le nombre déjà élevé d'interconnexions entre la couche d'étouffement et celle de traitement numérique de signaux. La seconde consiste à l'espace restreint du circuit d'étouffement, puisque chaque bit de configuration provenant de la couche de traitement numérique de signaux est alimenté à 1,5 V et doit par conséquent être augmenté à 3,3 V à l'aide d'un traducteur de niveau par bit, qui sera présenté à la section 3.2.9. La grande différence entre les deux monostables réside dans la grosseur des capacités utilisées. Les capacités du monostable de recharge sont beaucoup plus petites puisqu'elles doivent permettre un temps de recharge variable entre 10 ns et 100 ns environ, comparativement au monostable du temps mort qui doit permettre approximativement entre 10 ns et 1 µs.

3.2.8 Amorce et références de courant

Le circuit d'amorçage, plus communément appelé *bootstrap*, a pour but de procurer une référence de courant au reste du circuit. Il consiste en deux branches, dont les deux transistors PMOS M3 et M4 de la figure 3.12 forment un miroir de courant. Un équilibre se forme entre la tension grille-source du transistor NMOS M1 et celle développée aux bornes de la résistance R puisqu'elles sont connectées ensemble. Le point d'équilibre se trouve grâce à l'équation suivante [Allen et Holberg, 2002] :

$$I_2 R = V_{T1} + \left(\frac{2I_1 L_1}{K'_N W_1} \right)^{1/2} \quad (3.3)$$

Sachant que les courants I_1 et I_2 sont les mêmes en raison du miroir de courant, il est donc possible de trouver la valeur de ces courants au point d'équilibre en fonction des caractéristiques du transistor NMOS et de la résistance. Cependant, selon le graphique présenté à la figure 3.13, il existe deux points d'équilibre, soit un à la valeur de courant désiré et un second à l'état nul. Afin d'empêcher que le circuit ne converge vers le point d'équilibre mort, il est nécessaire de lui rajouter un petit module de démarrage. Les deux transistors NMOS M7 et M8, ainsi que la résistance à l'intérieur de l'encadré en pointillé de la figure 3.12 assurent qu'un petit courant circule par la branche 1, permettant au circuit de démarrer et d'atteindre le point d'équilibre désiré. Toutefois, ce petit circuit de démarrage consomme en permanence, ce qui lui confère un certain désavantage. L'alternative proposée pour éviter ce problème consiste en deux transistors PMOS branchés en diode, soit MP2 et MP3 de la figure 3.14. En étant placés tête-bêche, ils servent de grosse résistance contournant le transistor MN1, assurant ainsi une tension de masse au drain du

transistor PMOS MP1 et qu'un courant, si minime soit-il, circule dans la branche 2. Une fois le circuit démarré, l'impédance de ces transistors ainsi branchés dépasse largement celle du transistor MN1, rendant par le fait même le courant circulant par ces PMOS pratiquement nul. Sur le schéma de l'amorce conçue à la figure 3.14, la résistance R a été remplacée par 9 plus petites résistances disposées en série, en raison de sa trop grande valeur.

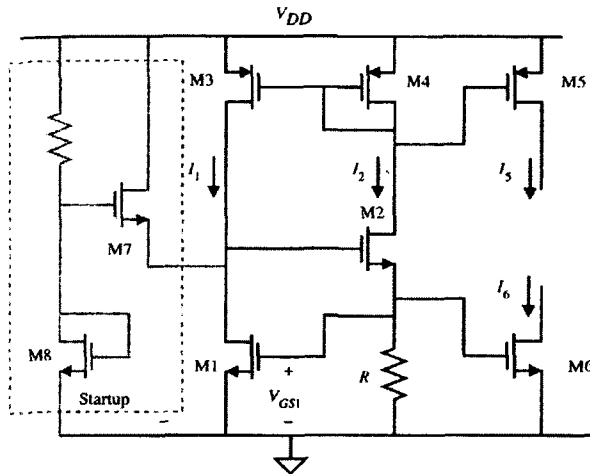


Figure 3.12 Amorce : Schéma [Allen et Holberg, 2002].

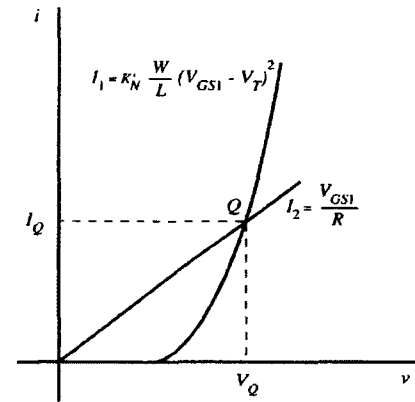


Figure 3.13 Amorce : Point d'équilibre [Allen et Holberg, 2002].

Les quatre dernières branches, à droite sur la figure 3.14, servent à fixer les courants de référence pour les différents monostables. Chacune des références est ajustée en modifiant la taille du transistor PMOS, changeant ainsi le ratio W/L et variant par la même occasion le coefficient multiplicateur de cette branche de référence. Sous les PMOS se retrouvent deux rangées de transistors NMOS, suivies d'un dernier NMOS, MN12. La seconde rangée de transistors NMOS sert de miroirs pour les références de courant des monostables. Ces transistors doivent se retrouver dans les mêmes conditions que les transistors des monostables, afin de s'assurer que les références de courant soient les plus identiques possible. C'est pourquoi la première rangée de transistors sous les PMOS sert à remplacer les transistors des bits de contrôle des monostables, alors que le transistor NMOS MN12 sous les quatre branches remplace le transistor NMOS de l'inverseur à l'entrée des monostables. Une seule référence de courant aurait pu être faite et avoir des transistors de différents ratios W/L à l'intérieur de chacun de monostables afin d'avoir chacune des références désirées. Ainsi, moins de références auraient été transférées d'un circuit à l'autre. Cependant, utiliser quatre références externes comme le cas présent permet d'avoir des transistors de tailles minimales à l'intérieur de chacun des monostables, réduisant la taille de ces mo-

dules à l'intérieur de chaque circuit d'étouffement pour aider à atteindre l'objectif de $50 \times 50 \mu\text{m}^2$ d'aire utilisée.

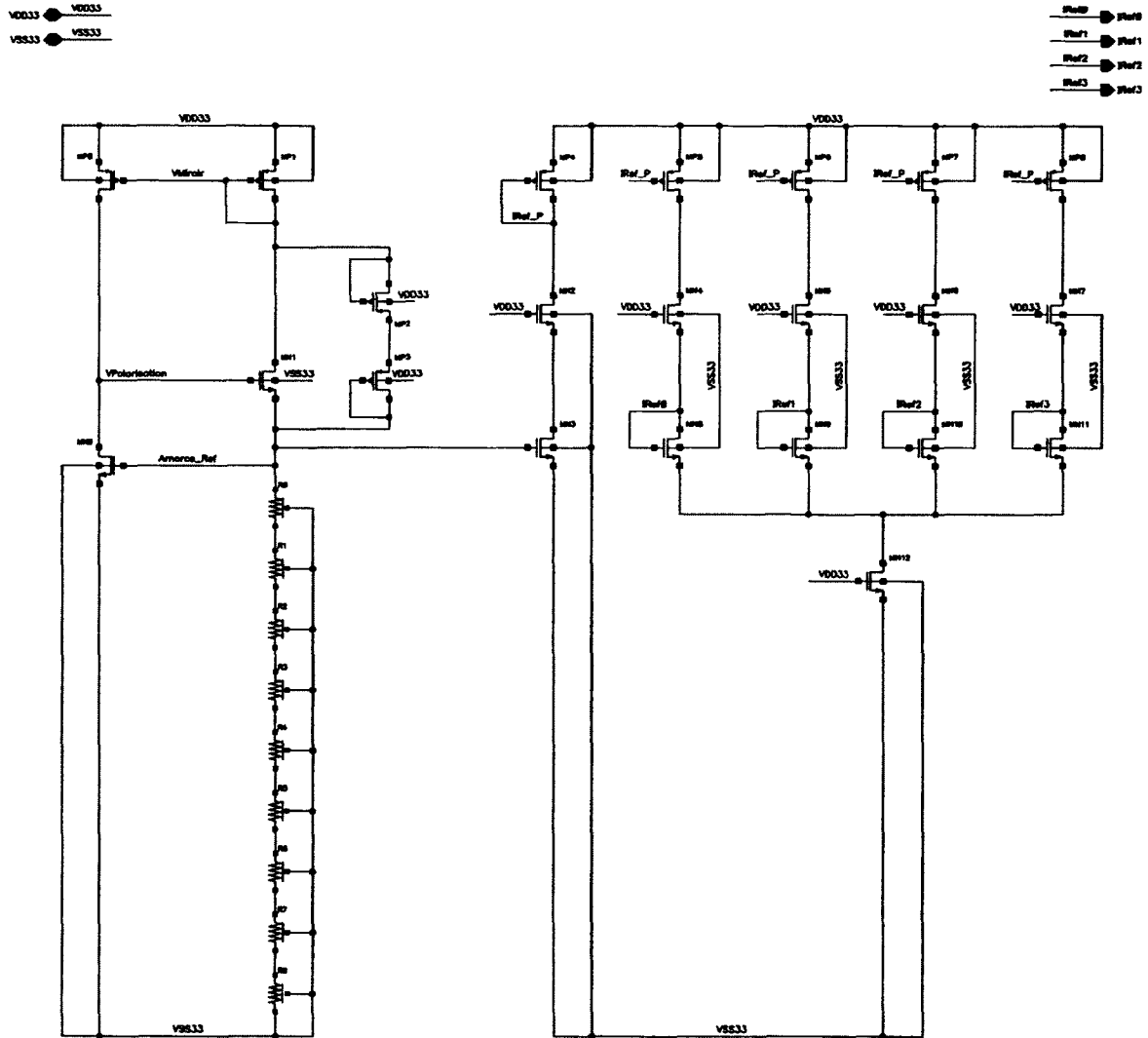


Figure 3.14 Schéma de l'amorce et références de courant.

3.2.9 Traducteur de niveau

Le traducteur de niveau permet le passage des signaux de l'échelle 0/1,5 V provenant de la couche de traitement numérique de signaux à une échelle 0/3,3 V, afin de se conformer à l'alimentation utilisée pour la couche d'étouffement. En effet, la couche de traitement numérique de signaux utilisant des transistors à oxyde mince, une alimentation de 1,5 V est requise, alors que la couche de circuits d'étouffement utilise des transistors à oxyde

épais, nécessitant une alimentation de 3,3 V. Le signal de 1,5 V et son inverse de 0 V se connectent sur chacune des grilles des deux transistors NMOS visibles à la figure 3.15. Le transistor NMOS MN1 activé par le signal 1,5 V laisse passer la tension de masse jusqu'à la grille du transistor PMOS MP0 de la branche opposée, laissant passer à son tour l'alimentation 3,3 V de la couche d'étouffement jusqu'au drain du transistor NMOS MN0 fermé par le signal inverse du bit de contrôle. Le signal de sortie 3,3 V et son inverse se retrouvent entre les drains des transistors PMOS et NMOS. Les transistors NMOS étant de type 3,3 V, mais alimentés seulement à 1,5 V, leur ratio W/L se doit d'être plus grand, réduisant ainsi leur résistivité afin de permettre la transition du circuit.

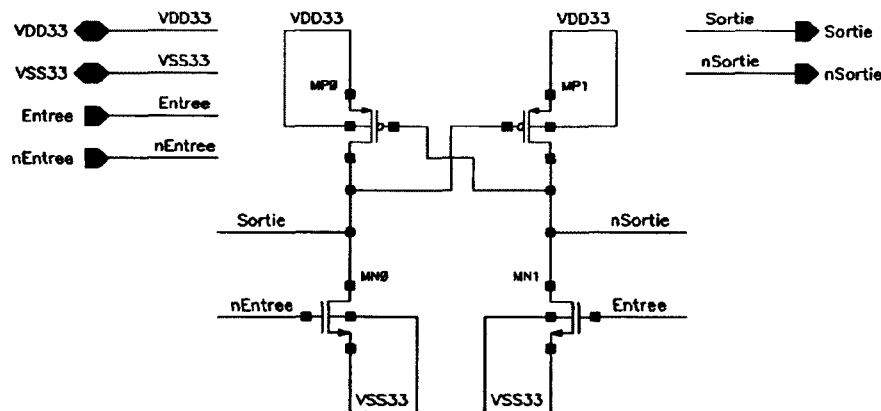


Figure 3.15 Schéma du traducteur de niveau.

3.2.10 Logique de contrôle

L'étage de logique de contrôle est divisé en trois parties : le contrôle du transistor de recharge de la PAMP, le contrôle du transistor de détection/étouffement, ainsi que le contrôle du signal de sortie. La première partie se compose d'une porte logique NON-OU (*NOR*), la seconde consiste en une porte logique ET (*AND*), alors que la troisième et dernière partie comporte deux portes logiques NON-ET (*NAND*). Le tout se retrouve à la figure 3.16. Les courbes des signaux présents à l'entrée de chacune des portes sont présentées à la figure 3.4. Afin de faciliter la compréhension de cette section, des diagrammes d'états de ces portes sont présentés à l'annexe C.

La porte logique NON-OU du contrôle de la recharge de la photodiode a besoin que les deux entrées soient au niveau bas pour obtenir un niveau haut à la sortie. Le premier signal d'entrée, soit *MsContrôle* provenant de la sortie de la bascule, est de niveau préalablement bas et prend un niveau haut pour la durée du temps mort avant de revenir à son état initial. Le second signal d'entrée, soit *TempsRecharge* provenant de la sortie du monostable du

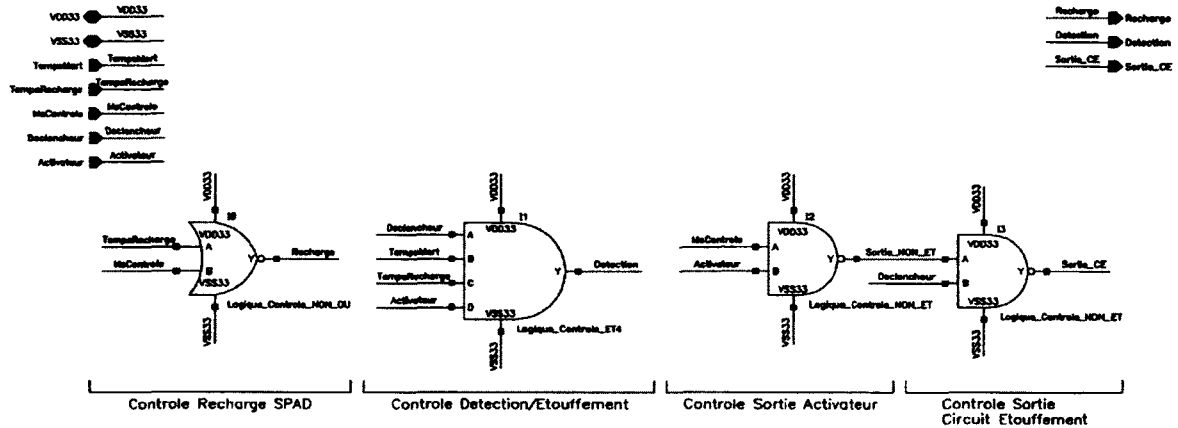


Figure 3.16 Schéma-bloc de la logique de contrôle.

temps de recharge, est de niveau haut et prend un niveau bas pour la durée du temps mort et du temps de recharge avant de revenir à son état initial. Les deux entrées ne sont donc toutes deux de niveau bas que pour la période déterminée par le monostable de recharge, activant ainsi le transistor NMOS de recharge à l'entrée du circuit pour cette durée. Du côté de la conception de la porte, dont la connexion des transistors est montrée à la figure 3.17, il est important que les ratios W/L des transistors permettent une conduction du courant au moins équivalente à un inverseur régulier. La pire conduction possible pour les NMOS survient lorsqu'un seul transistor n'est actif, ne nécessitant par conséquent rien de plus que le ratio W/L habituel. Lorsque la branche des transistors PMOS conduit, comme les transistors sont branchés en série, les impédances équivalentes s'additionnent. Pour la présente porte, avec deux transistors PMOS en série, leurs ratios W/L doivent être deux fois plus grands.

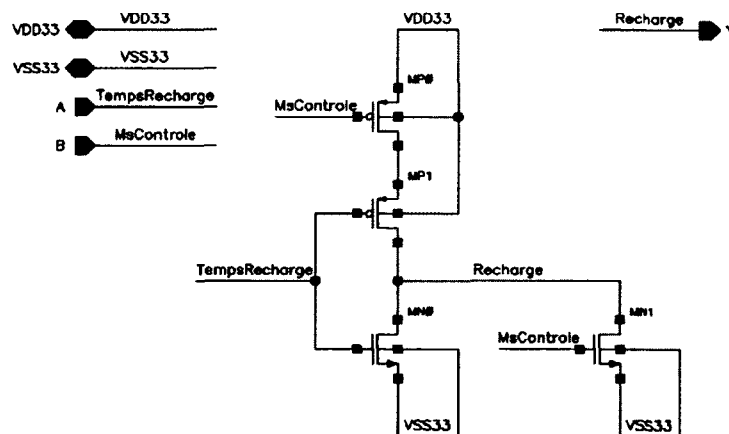


Figure 3.17 Schéma de la porte logique NON-OU.

Le contrôle du transistor de détection/étouffement possède quatre entrées, soient le signal de sortie du comparateur de tension (*Declencheur*), le signal de sortie du monostable du temps mort (*TempsMort*), le signal de sortie du monostable du temps de recharge (*TempsRecharge*), ainsi que le bit de contrôle activant le circuit d'étouffement (*Activateur*), comme démontrés à la figure 3.18. Ce dernier provenant de la couche de traitement numérique de signaux, le signal doit par conséquent être mis à niveau par le traducteur présenté à la section précédente. À l'état initial, tous les signaux à l'entrée de la porte sont de niveau haut, reproduisant un niveau haut à sa sortie et activant par le fait même le transistor de détection et d'étouffement. Lors de la détection d'un événement, la sortie de la porte change d'état pour adopter un niveau bas. La seconde entrée maintient la sortie à ce niveau pour la durée du temps mort, alors que la troisième fait de même pour le temps de la recharge. Une fois le temps de recharge terminé, toutes les entrées sont de nouveau à un niveau haut et le transistor de détection/étouffement est réactivé. Lorsque le bit de contrôle activant le circuit se trouve à un niveau bas, la sortie de la porte logique est forcée à un niveau bas et éteint la branche de détection. Le circuit ne pouvant faire de détection, la branche de recharge ne peut jamais s'activer, isolant ainsi l'anode de la photodiode. Par conséquent, dès que la photodiode se déclenche, elle se retrouve étouffée sur le champ et demeure dans cet état sans être rechargée.

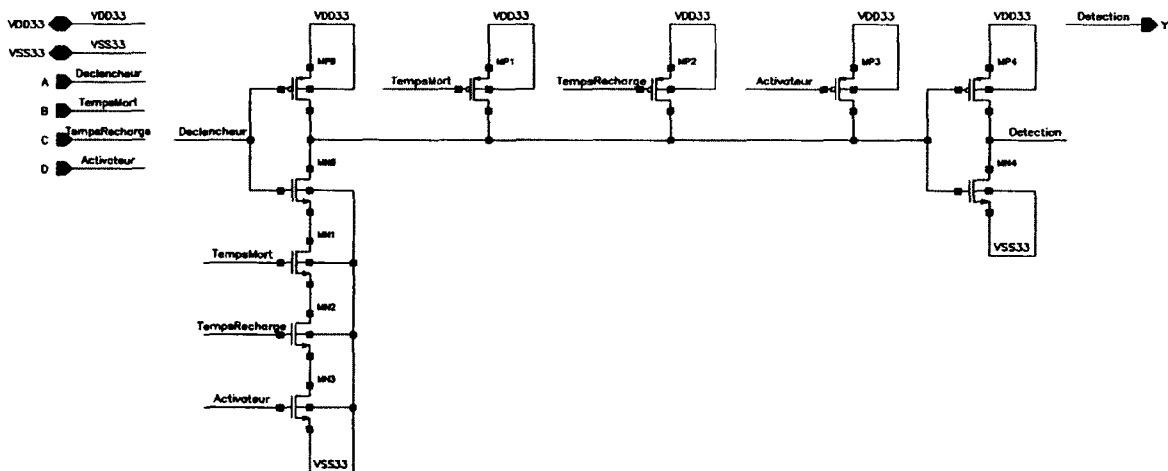


Figure 3.18 Schéma de la porte logique ET.

Pour le contrôle du signal de sortie, chacune des deux portes logiques NON-ET possède deux entrées, telles que présentées à la figure 3.19. La première utilise la sortie de la bascule (*MsContrôle*), ainsi que le bit de contrôle activant le circuit (*Activateur*). Le signal provenant de la bascule permet d'obtenir un signal de sortie du délai du temps mort. Ce choix de conception provient des limitations de la couche de traitement numérique de

signaux. En effet, la fréquence de l'horloge utilisée étant de 200 MHz, le signal de sortie du circuit d'étouffement doit donc revenir à un niveau bas pour un minimum de 5 ns entre deux événements, afin que la logique numérique puisse les différencier. Advenant le cas où un photon serait absorbé pendant la recharge de la photodiode, le circuit serait déclenché à nouveau dès la recharge terminée. Dans ce cas, si le signal de sortie était tenu à un niveau haut pour toute la durée où le circuit est insensible, soit la durée du temps mort ainsi que celle du temps de recharge, la logique numérique ne pourrait différencier les deux événements. Le temps de recharge étant d'au moins 10 ns, en utilisant un signal de sortie de la durée du temps mort, cela assure d'avoir un signal de sortie bas de plus de 5 ns et de compter deux événements advenant le cas d'un nouveau déclenchement de la photodiode pendant la recharge. De plus, le bit de contrôle assure de garder la sortie du circuit d'étouffement à un niveau bas dans le cas où le circuit serait désactivé. Malgré le fait que le circuit soit désactivé, il pourrait subvenir le cas où le bruit à l'entrée du circuit soit assez fort pour faire déclencher le circuit. Cette configuration préserve un niveau bas à la sortie et garantit ainsi que la logique numérique ne considère pas ce déclenchement erroné. La sortie de cette première porte donne un niveau logique bas en activation et sert d'entrée pour la seconde porte logique. La seconde entrée provient de la sortie du comparateur de tension à l'entrée du circuit d'étouffement (*Declencheur*). Cette porte a deux utilités, la première étant d'inverser le niveau du signal de sortie afin d'obtenir un niveau haut lors de la détection. La seconde est l'obtention d'un changement d'état de la sortie plus rapide et avec moins de gigue en utilisant directement la sortie du comparateur de tension de l'entrée du circuit d'étouffement comme déclencheur.

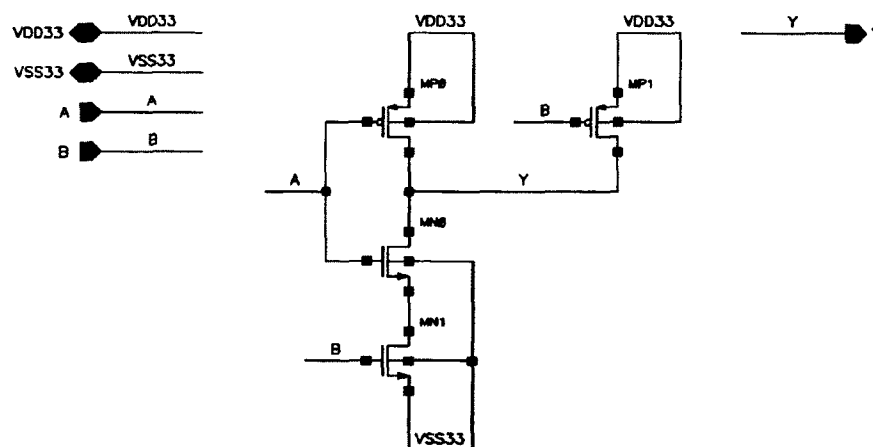


Figure 3.19 Schéma de la porte logique NON-ET.

3.2.11 Générateur d'impulsions

Le générateur d'impulsions ne requiert que trois transistors pour sa conception, soit un transistor PMOS branché en série avec un transistor NMOS, avec la grille d'un second transistor NMOS rattachée entre les deux (figure 3.20). Le transistor PMOS MP0 sert à la recharge du générateur, alors que le transistor NMOS MN0 sert à la décharge. Le transistor NMOS MN1 exploite quant à lui sa capacité de grille, afin d'accumuler des charges servant à déclencher le circuit par la suite.

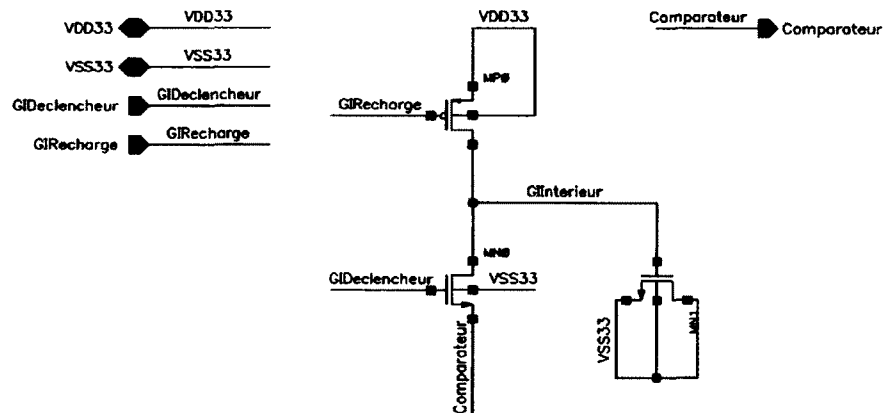


Figure 3.20 Schéma du générateur d'impulsions.

À l'état initial, le transistor NMOS MN0 de décharge est éteint. Le transistor PMOS MP0 est alors allumé pendant un coup d'horloge par le signal *GIRecharge*, permettant de charger la capacité MOS à la tension d'alimentation, soit 3,3 V, par le nœud *GIInterieur*. Le transistor de recharge est ensuite éteint et le générateur attend une impulsion. La logique numérique active ensuite le transistor MN0 pendant un coup d'horloge avec le signal *GIDeclencheur*, déchargeant ainsi la capacité à travers le transistor MN2 à l'entrée du comparateur de tension de la figure 3.6. La capacité étant chargée à la tension d'alimentation, cela occasionne une impulsion de courant lorsqu'elle se fait forcer à la tension de masse. Cet effet réplique sensiblement le même comportement que lorsque la photodiode développe son avalanche de courant. Les charges circulant dans le transistor NMOS MN2 de l'étage primaire du circuit d'étouffement développent une tension assez grande à l'entrée du comparateur de tension pour déclencher celui-ci. Lors du coup d'horloge suivant, la couche de traitement numérique de signaux désactive le transistor de décharge MN0 et active celui de recharge MP0 afin de préparer le générateur pour l'impulsion suivante. Les simulations, montrées à la figure 3.21, ont permis d'ajuster les deux transistors NMOS afin de reproduire environ le même nombre de charges et la même vitesse de décharge que le comportement simulé par le modèle de la PAMP. En effet, cette figure montre un

temps de montée de la tension à l'entrée du comparateur semblable entre la simulation avec la PAMP et celle avec le générateur d'impulsions. La tension plus élevée dans le cas du générateur d'impulsions s'explique par le fait que ce dernier se connecte directement au nœud à l'entrée du comparateur et que cette tension se stabilise par le diviseur de tension formé par le transistor de décharge MN0 et le transistor MN2 de l'étage primaire du circuit d'étouffement. Pour la simulation avec une PAMP, dès que l'événement est détecté, le transistor d'étouffement est éteint, empêchant la tension d'augmenter à l'entrée du comparateur. Les simulations montrent également que le signal *Declencheur* à la sortie du comparateur détecte l'événement sensiblement en même temps et que les courbes des signaux de sortie se superposent pratiquement parfaitement. Une fois ces deux composants ajustés, la taille du transistor PMOS a été réglée pour assurer la recharge de la capacité en moins d'un coup d'horloge.

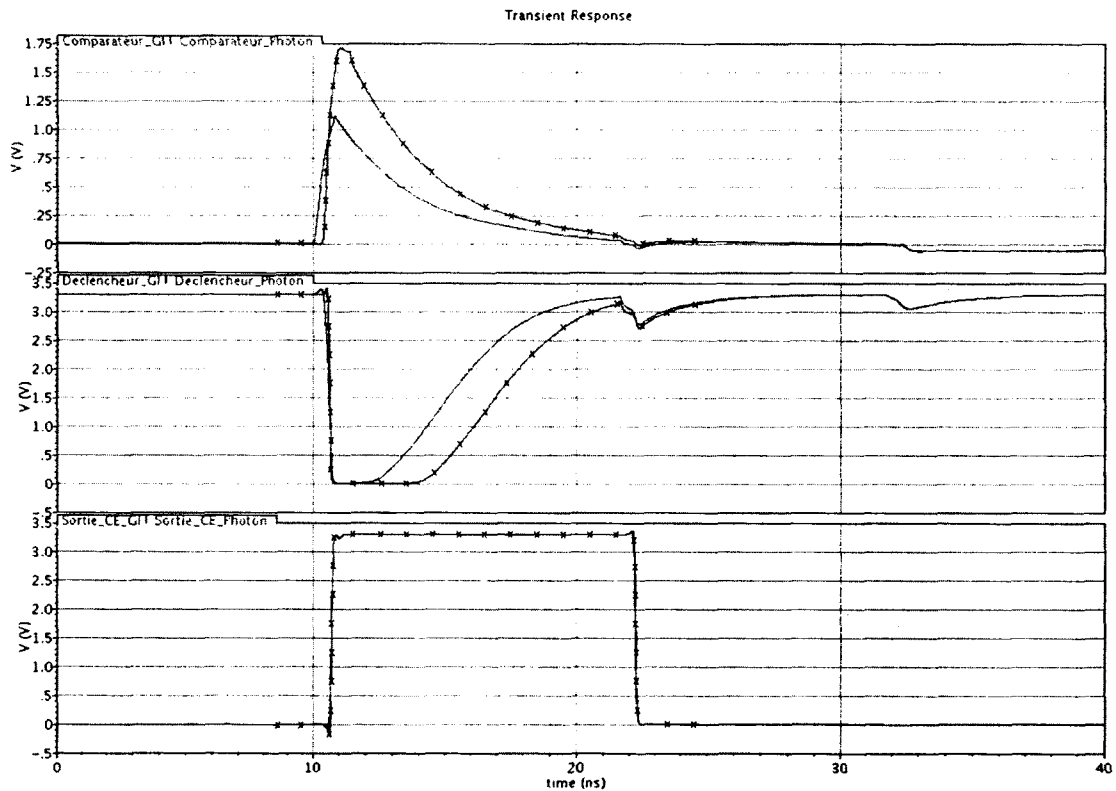


Figure 3.21 Comparaison entre les simulations du logiciel *Cadence* avec la PAMP et avec le générateur d'impulsions. Les courbes rouges proviennent de la simulation avec la PAMP. Les courbes bleues avec les « x » proviennent de la simulation avec le générateur d'impulsions.

3.2.12 Simulations de matrices

Une fois tous ces différents modules conçus, il est primordial de faire une simulation d'une matrice complète. La simulation comporte 484 fois le circuit d'étouffement, répartie en 22×22 circuits, ainsi qu'une amorce et références de courant pour tous les monostables de la matrice complète. Il en est de même pour les traducteurs de niveaux des bits de contrôle des monostables de recharge ; ces traducteurs de niveaux ne sont présents qu'une seule fois par matrice, puisque tous les monostables de recharge de la matrice utilisent les mêmes quatre bits.

Considérant le nombre élevé de circuits d'étouffement présents sur la matrice complète, ainsi que le nombre imposant d'interconnexions s'y rattachant, le temps requis pour faire cette schématisation à la main est impraticable. La solution réside en un fichier de description écrit en langage de programmation VHDL. Ce fichier fait l'énumération de tous les modules se trouvant dans la matrice. Le logiciel *Synopsys* joue le rôle d'intermédiaire et rend le fichier compréhensible pour le logiciel *Cadence*, c'est-à-dire qu'il transpose le code VHDL en portes logiques. Par exemple, pour produire un schéma contenant une matrice de circuits d'étouffement, il suffit simplement de décrire un circuit d'étouffement par son nom, ses entrées et sa sortie dans le fichier VHDL et de lui faire subir une boucle qui répétera le circuit jusqu'à en obtenir 484. Le nom des instances, des bits de configuration, des sorties, etc. sont incrémentés automatiquement et les 484 circuits sont correctement interconnectés dans *Cadence*, et ce, après seulement quelques lignes de codes, ainsi qu'une compilation dans *Synopsys*. Le même procédé a été utilisé pour former les 484 photodiodes se connectant chacune à son circuit d'étouffement.

Deux autres modules ont été conçus pour les simulations, soit un premier pour contrôler les photodiodes qui se déclenchent et un second pour contrôler les bits de configuration se rendant à chacun des circuits d'étouffement. Tous deux se retrouvent sous le langage de programmation Verilog-A. Le premier module permet quatre options : aucune PAMP ne se déclenche, toutes les PAMP se déclenchent, toutes les PAMP se déclenchent à l'exception de certaines photodiodes spécifiées ou aucune PAMP ne se déclenche à l'exception de certaines photodiodes spécifiées. Le second module prend chacun des bits de configuration et le transmet à tous les circuits d'étouffement. La même configuration est donc utilisée pour tous les circuits de la matrice pour la simulation et ce module permet un changement de configuration rapide pour tous les circuits.

3.2.13 Schéma de la couche complète

Le schéma de la couche complète permet tout d'abord de s'assurer que les numéros de bits donnés par la couche de traitement numérique de signaux fournissent les bons circuits sur la couche d'étouffement. Cela permet également, une fois la conception des masques faite, de faire une vérification entre les dessins de masques et les schématiques (*Layout Vs Schematic*, LVS) pour s'assurer de la connectivité des signaux.

Encore une fois, le nombre élevé de composants présents sur la couche complète, soit 2904 excluant les circuits de tests, impose trop d'interconnexions pour schématiser la couche à la main. La même méthode que pour les simulations par matrice a été utilisée, soit de programmer la couche par code VHDL et de passer ensuite par le logiciel *Synopsys* avant d'importer le résultat dans *Cadence*. Aucune simulation n'a cependant pu être faite pour la couche complète, la simulation étant trop lourde et le temps requis trop important.

3.3 Conception des masques

3.3.1 Masques du circuit et mise en matrice

Une fois les simulations complétées et les tailles des composants choisies, l'étape suivante consiste à faire la conception des masques afin de rendre possible la fabrication des circuits. Les fichiers GDS sont préparés et envoyés à la fonderie, afin que les compagnies Chartered et Tezzaron Semiconductor fabriquent la puce. Les choix de conception pour la réalisation des masques des circuits contribueront particulièrement à l'espace utilisé et, ainsi, à la réussite ou l'échec de l'objectif de dimensions du circuit de $50 \times 50 \mu\text{m}^2$. L'élaboration de ces masques est primordiale puisqu'elle influence grandement, entre autres, les capacités parasites et la diaphonie électromagnétique à l'intérieur du circuit, mais encore plus directement la capacité parasite à l'anode de la photodiode et, par conséquent, les performances de cette dernière. C'est pourquoi les transistors d'entrée se retrouvent juxtaposés aux TSV. Afin d'assurer un temps de montée rapide lorsqu'il est déclenché, le générateur d'impulsions se doit également d'être près de l'électronique frontale réduisant la capacité parasite entre ces deux modules.

Les dessins des masques ont été faits un module à la fois, tels qu'ils ont été présentés à la section 3.2. Cette façon de faire permet de vérifier la bonne fonctionnalité, ainsi que les effets du dessin sur le circuit, module par module. Une fois le dessin des masques terminé pour un module, la première étape de vérification consiste à faire une comparaison entre ce dessin et le schématique (*Layout Vs Schematic*, LVS). Cela vient confirmer que tous les

composants présents dans le schématique ont été dessinés tels qu'ils y sont décrits et, surtout, qu'ils sont correctement connectés entre eux. Une extraction parasitique sert ensuite, comme son nom l'indique, à extraire le circuit dessiné tout en considérant l'influence et les conséquences de l'architecture du dessin ; il ajoute, entre autres, la résistance des traces en fonction de leur longueur et de leur largeur, ainsi que les capacités parasites entre les traces et les composants du circuit. Une nouvelle simulation avec ces paramètres permet d'observer les effets parasites sur le circuit dépendamment de la façon dont le circuit a été dessiné et d'y apporter des changements et améliorations en conséquence. Le dessin des masques du circuit complet est présenté à la figure 3.22, sur laquelle le module de la bascule a été inclus dans le bloc de l'électronique frontale. Le métal à l'arrière de la couche et le métal 6 ne sont pas présents sur la figure afin de mieux voir les différents modules du circuit.

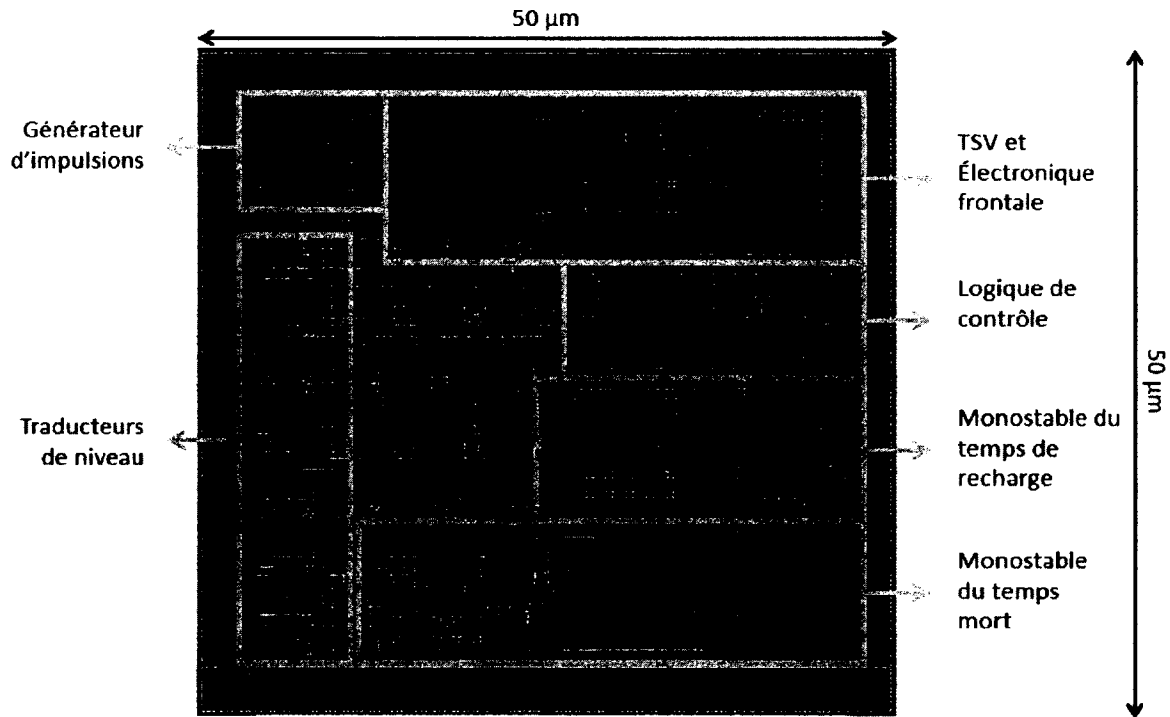


Figure 3.22 Dessin des masques du circuit d'étouffement avec des transistors 3,3 V en entrée.

Afin d'assurer une bonne polarisation du substrat, tout en prévenant le déclenchement parasitique (*latch-up*), un anneau de contacts au substrat entoure chacun des modules. Ces anneaux se composent de contacts de même taille et espacés de la distance minimale requise par le fabricant. Ainsi, il est facile de juxtaposer les modules les uns aux autres, les contacts des anneaux de garde se superposant parfaitement.

Les références de courant (figures 3.9, 3.11, 3.14) devant être les mêmes d'un circuit à l'autre, cette partie de l'électronique requiert un appariement. C'est pourquoi les quatre branches de référence sont identiques, de forme et d'orientation, d'un monostable à l'autre, ainsi qu'avec le circuit de référence. Comme les branches pour IRef0, IRef1, IRef2 et IRef3 se retrouvent individuellement dans le même environnement dans le circuit de référence et dans les monostables, l'ajout de faux éléments (*dummy*) a été jugé facultatif pour cet appariement. Considérant la contrainte d'espace, cette option d'ajout n'a par conséquent pas été retenue.

Les circuits d'étouffement étant matricés en 22×22 circuits tels que montrés à la figure 3.23, il est crucial d'avoir un bon maillage d'alimentation afin d'éviter le plus possible les chutes de tension dues à la résistivité des traces. En couvrant l'espace inutilisé en périphérie du circuit, on obtient approximativement $4 \mu\text{m}$ de chaque côté du circuit, permettant une largeur de trace de $8 \mu\text{m}$ entre chaque circuit lorsqu'ils sont matricés. Le procédé de fabrication rend disponibles six étages de métaux pour la conception du circuit intégré, le dernier étage étant réservé pour faire les interconnexions face à face entre les couches. Il demeure donc cinq étages pouvant servir à la circulation de l'alimentation et des signaux. Les trois premiers étages de métaux étant plus résistifs que les deux du haut, les trois premières couches ont été assignées à la masse et les deux dernières pour l'alimentation positive. Afin de faciliter l'acheminement des signaux à l'intérieur d'un circuit, les métaux 2 et 4 permettent la circulation des signaux verticalement, alors que les métaux 3 et 5 les font voyager horizontalement. Cette technique facilite le dessin des traces et assure qu'aucun court-circuit ne soit fait entre deux traces sur le même niveau, puisque toutes les traces d'un même niveau sont parallèles les unes aux autres. Les signaux de références de courant, ainsi que les bits pour le temps de recharge doivent se propager et atteindre tous les circuits d'une même matrice. Pour permettre leur acheminement, ces signaux partagent les $8 \mu\text{m}$ réservés pour l'alimentation, en utilisant la même méthodologie que celle expliquée précédemment pour la circulation verticale et horizontale des signaux. Ces signaux étant envoyés aux circuits avant le début de l'acquisition de données et ne changeant pas d'état pendant l'acquisition, leurs délais de transmission ne sont pas critiques. Le module de références de courant et les traducteurs de niveaux pour les bits des monostables de recharge, en étant les mêmes pour toute la matrice, ont été placés dans une partie de l'espace mort de $50 \mu\text{m}$ sous l'isolant du scintillateur, ne dérangeant ainsi aucunement l'architecture de la matrice, tels qu'ils sont indiqués à la figure 3.23.

Finalement, l'espace inutilisé à l'intérieur d'un circuit d'étouffement a été rempli par des capacités MOS. Ces capacités permettront d'aider à stabiliser l'alimentation en fournis-

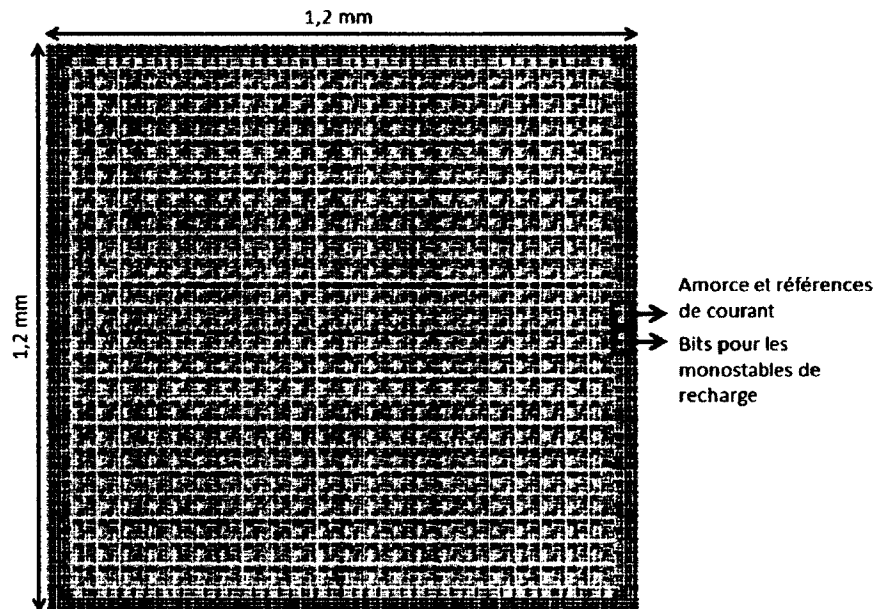


Figure 3.23 Dessin des masques d'une matrice de circuits d'étouffement.

sant rapidement quelques charges et ainsi réduire les chutes de tension. Tel que mentionné précédemment, le dernier niveau de métal sert à faire les interconnexions entre la couche d'étouffement et celle de traitement numérique des signaux par l'intermédiaire d'octogones. Une redondance a été faite afin de s'assurer que les signaux puissent se rendre sans problème d'une couche à l'autre, même s'il advenait un problème avec un plot lors de la fabrication, permettant donc à chaque signal de passer par deux plots d'interconnexion.

3.3.2 Masques des DEMOS et des matrices à haute tension

Tel qu'expliqué à la section 3.2.4, les simulations effectuées avec les modèles des transistors 5 V du logiciel *Cadence* ont permis de donner une bonne approximation des tailles nécessaires pour les transistors DEMOS, réalisés selon la recette présentée dans l'état de l'art à la section 2.3. Les restrictions en espace ont quelque peu limité les tailles de ces transistors et ne permettaient pas une pleine liberté quant à leur conception. Cependant, lors du dessin des masques du circuit d'étouffement avec les transistors 3,3 V (section 3.3.1), un espace excédentaire près des TSV a été comblé par des capacités MOS. Cet espace a été positionné près de l'entrée du circuit en prévision d'utiliser cette aire pour les transistors DEMOS, donnant ainsi un peu plus de flexibilité quant à la conception des DEMOS. Grâce aux tailles approximées par les simulations, les DEMOS ont pu entrer de justesse dans cet espace réservé du circuit d'étouffement (figure 3.24).

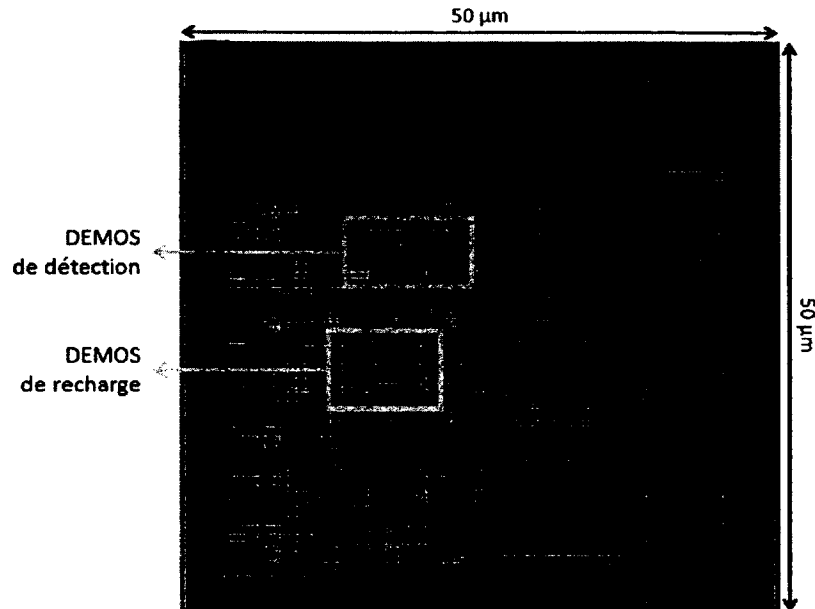


Figure 3.24 Dessin des masques du circuit d'étouffement avec des transistors DEMOS en entrée avec les mêmes modules que la figure 3.22.

3.3.3 Circuits de tests

La figure 3.25 présente le dessin des masques de la couche complète d'étouffement. Six matrices de 484 circuits d'étouffement chacune s'y retrouvent, dont quatre d'entre elles ont les transistors 3,3 V en entrée, alors que les deux autres ont les transistors DEMOS à la place. En plus des matrices de circuits d'étouffement, plusieurs structures de tests ont été mises en place sur la couche d'étouffement.

Tout d'abord, sur deux circuits des matrices avec les transistors 3,3 V en entrée, ainsi que pour deux circuits des matrices avec les transistors DEMOS en entrée, différents points de tests ont été inclus. Ces points de test ont été placés dans l'optique de vérifier la fonctionnalité des différents modules des circuits d'étouffement. Huit nœuds sont examinés :

- *Compareur* ;
- *Declencheur* ;
- *MsFin* ;
- *MsControle* ;
- *TempsRecharge* ;
- *Recharge* ;

- *Detection* ;
- *Sortie_CE*.

Les signaux *Comparateur* et *Declencheur* étant trop faibles, ils ont dû être amplifiés par une série d'inverseurs avant d'être sortis sur un plot d'interconnexion analogique. Pour faire la conception de cette série d'inverseurs, la largeur de canal W de chacun des deux transistors est augmentée progressivement du ratio A :

$$A = \left(\frac{C_{charge}}{C_{entrée}} \right)^{\frac{1}{N}} \quad (3.4)$$

où N , le nombre d'inverseurs nécessaires, se trouve par :

$$N = \ln \left(\frac{C_{charge}}{C_{entrée}} \right) \quad (3.5)$$

Ensuite, afin de calculer la gigue du circuit, deux circuits d'étouffement de chacun des deux types ont été insérés avec leur entrée et leur sortie entre deux plots d'interconnexion, un de ces circuits étant présenté en exemple à la figure 3.26. Le premier circuit de chaque type connecte son entrée directement au plot d'interconnexion, alors que le second possède une capacité en série entre son entrée et le plot. Cela a été fait afin d'augmenter la testabilité. Le signal de sortie étant trop faible encore une fois pour être mesuré directement par une sonde, un tampon de sortie a été utilisé, soit la série d'inverseurs mentionnée précédemment. Afin de donner plus de flexibilité aux tests, ces deux circuits ont été répétés intégralement à l'exception de l'ajout d'une capacité interne en série à l'entrée du circuit.

Les monostables de temps mort et de recharge possèdent également leurs structures de tests, montrées à la figure 3.27. Les deux modules ont été placés avec leur entrée et leur sortie connectées à un plot d'interconnexion. Cependant, l'inverseur à la sortie du monostable de temps mort, tout comme l'inverseur à l'entrée du monostable du temps de recharge, a été omis pour se connecter à l'inverseur d'entrée et à la sortie de la bascule de Schmitt dans les deux cas. Les différents temps morts attendus sont présentés au tableau 3.1, alors que ceux du temps de recharge se retrouvent au tableau 3.2. Il est à noter que le temps mort réel de la photodiode sera moindre que ceux présentés dans le tableau 3.1. En effet, le monostable commence à calculer son délai dès la détection d'une avalanche, soit avant même que la photodiode ne soit étouffée.

3.3. CONCEPTION DES MASQUES

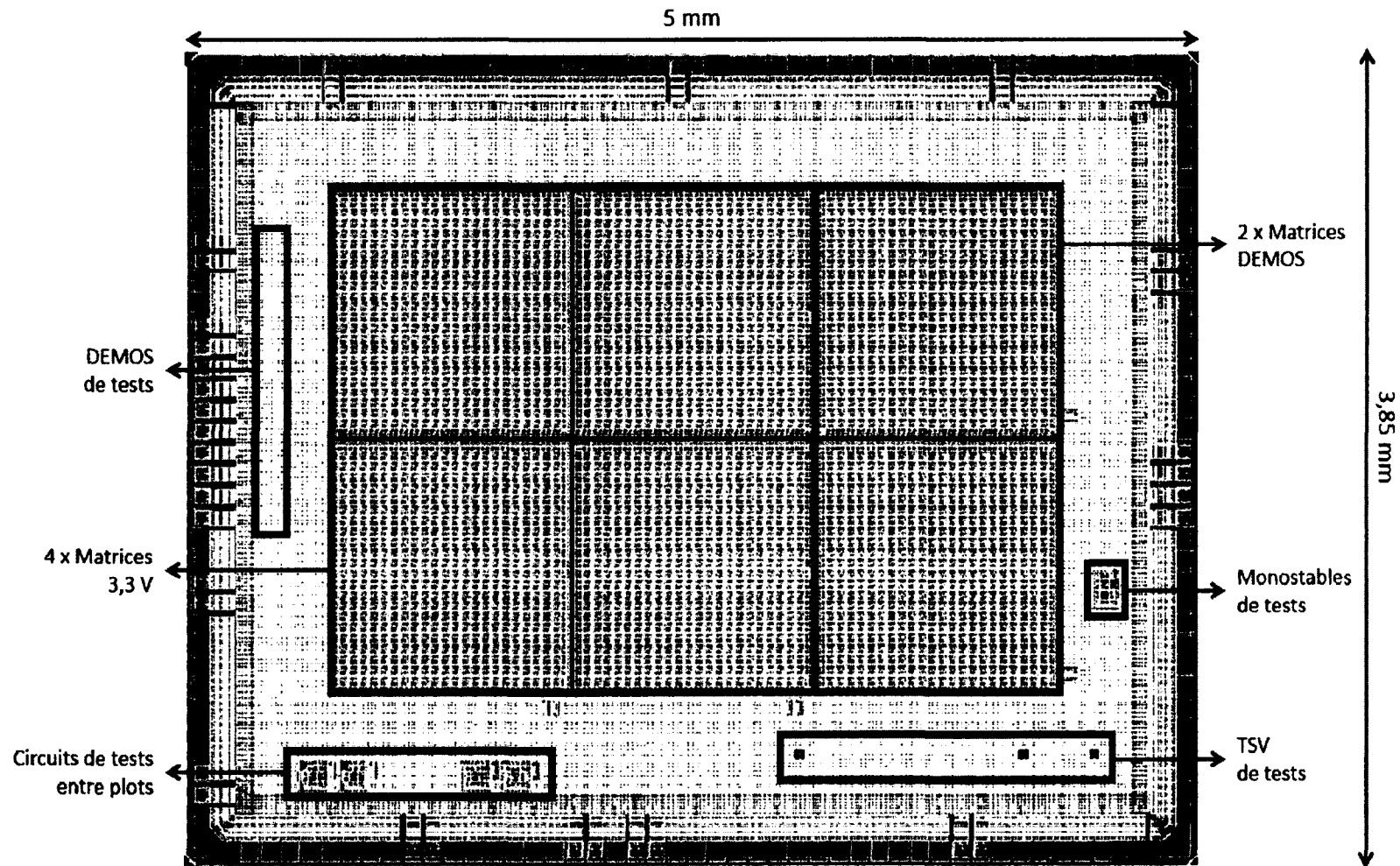


Figure 3.25 Dessin des masques de la couche complète d'étouffement.

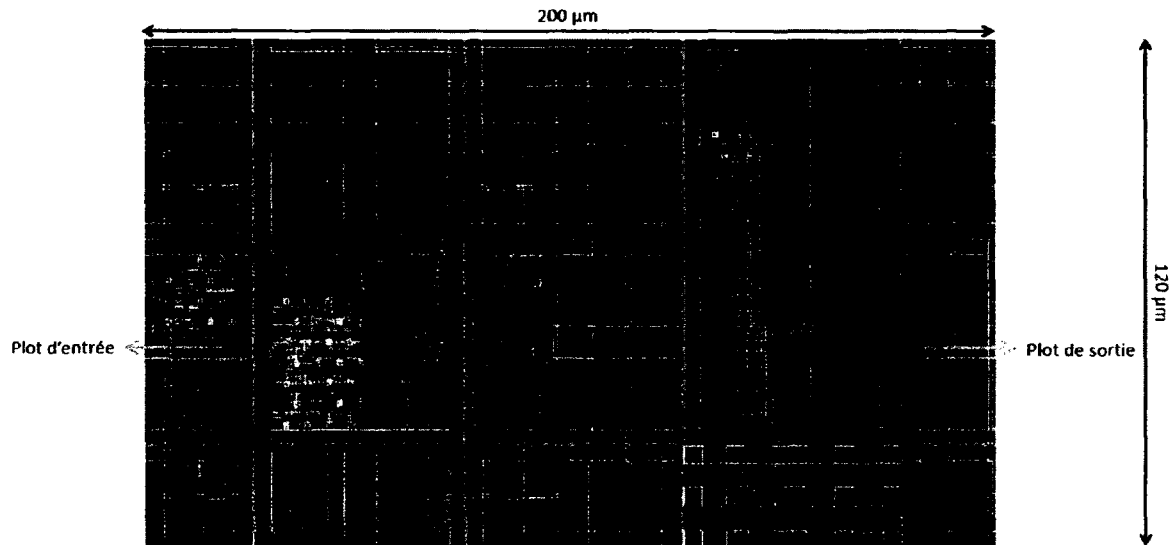


Figure 3.26 Dessin des masques d'un circuit d'étoffement de tests placé entre deux plots d'interconnexion.

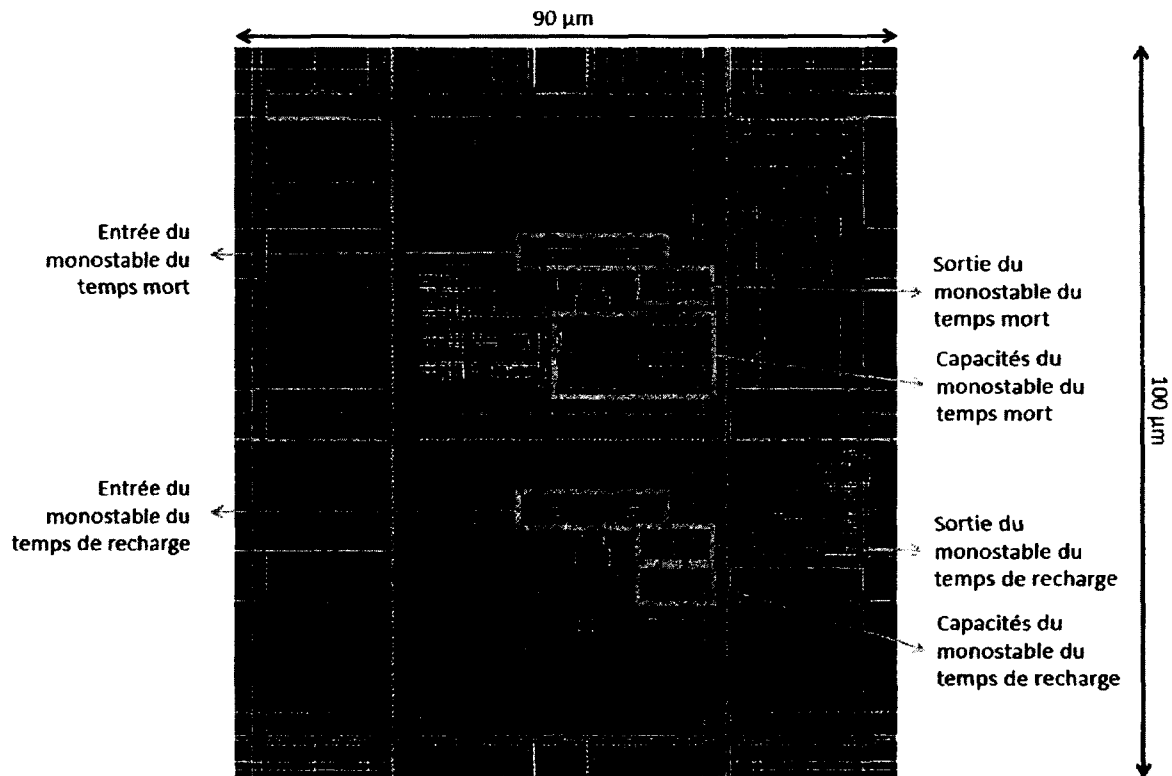


Figure 3.27 Dessin des masques des monostables de tests.

En plus des deux transistors DEMOS utilisés pour les circuits d'étouffement (transistors DEMOS 1 et 2), sept autres transistors DEMOS avec des paramètres différents ont été ajoutés pour des tests. Il sera ainsi possible de mieux étudier l'influence de la variation de chacun des paramètres vus à la section sur la conception de transistors DEMOS. Le tableau 3.3 présente les paramètres de chacun des transistors conçus, où L est la longueur du canal, L_a est la longueur de la région d'accumulation et L_d est la longueur de la région de dérive, telles que montrées à la figure 3.28.

Tableau 3.3 Dimensions des transistors DEMOS.

| DEMOS | L (μm) | L_a (μm) | L_d (μm) | W (μm) |
|-----------|-----------------------|-------------------------|-------------------------|-----------------------|
| DEMOS 0 | 4,0 | 2,0 | 4,0 | 2,0 |
| DEMOS 1* | 1,0 | 0,5 | 2,0 | 2,0 |
| DEMOS 2** | 1,0 | 0,5 | 2,0 | 0,7 |
| DEMOS 3 | 1,0 | 0,5 | 4,0 | 2,0 |
| DEMOS 4 | 1,0 | 2,0 | 2,0 | 2,0 |
| DEMOS 5 | 1,0 | 2,0 | 4,0 | 2,0 |
| DEMOS 6 | 4,0 | 0,5 | 2,0 | 2,0 |
| DEMOS 7 | 4,0 | 0,5 | 4,0 | 2,0 |
| DEMOS 8 | 4,0 | 2,0 | 2,0 | 2,0 |

* Transistor de recharge

** Transistor de détection/étouffement

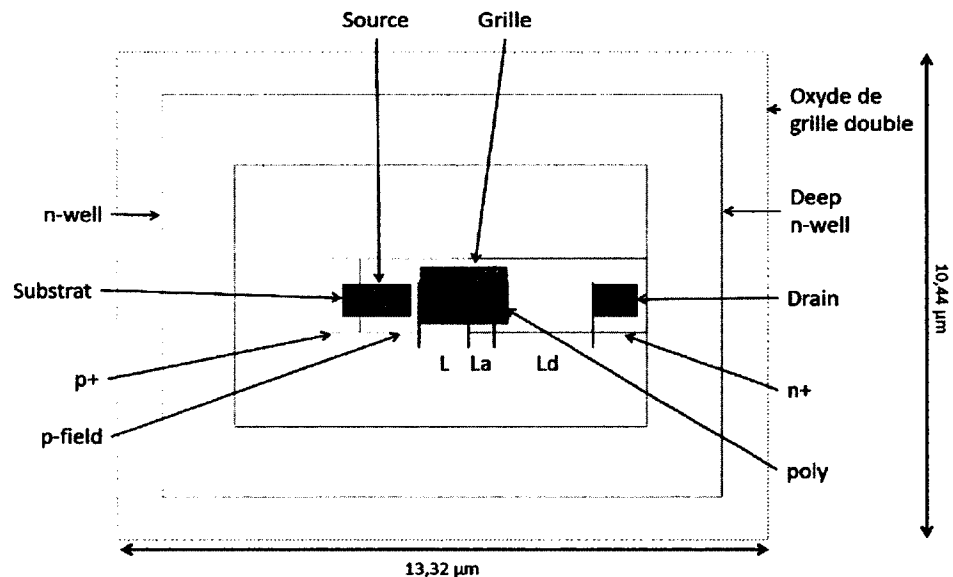


Figure 3.28 Dessin des masques d'un transistor DEMOS.

Finalement, dans l'optique d'une version future de cette puce où la haute tension pour alimenter les PAMP passerait directement au travers de la puce, trois structures pour

vérifier la haute tension sur les TSV ont été ajoutées. Les structures n'étant accessibles que par le derrière de la puce, chacune d'elles se connecte par conséquent par deux TSV. La première structure de test amène la haute tension par un TSV et monte jusqu'au métal 1. Le second TSV amène la masse à un anneau autour du premier TSV s'arrêtant au métal 1 également, avec une distance, représentée par la variable d à la figure 3.29, d'environ $14,6 \mu\text{m}$ entre les deux métaux. La seconde structure (figure 3.30) se rend jusqu'à des plots d'interconnexion de métal 6 éloignés d'environ $11 \mu\text{m}$. La dernière structure (figure 3.31) se rend également jusqu'aux plots d'interconnexion du sixième métal, mais cette fois sur deux plots directement adjacents.

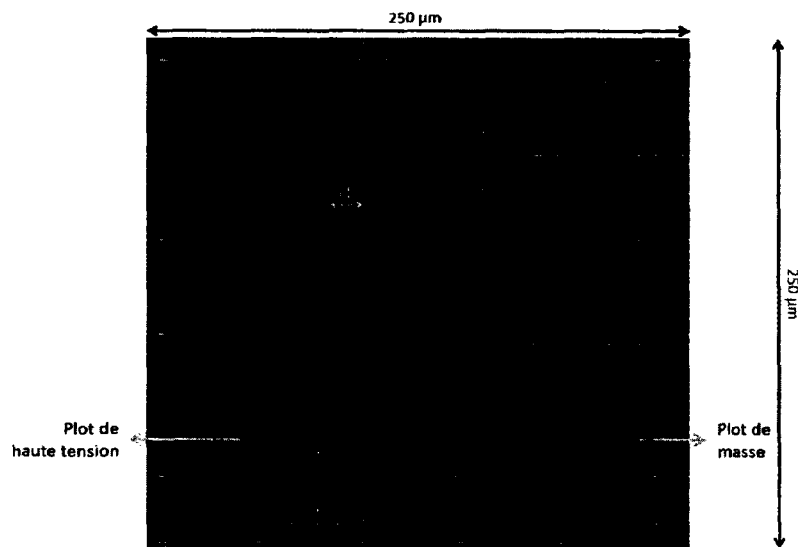


Figure 3.29 Dessin des masques du premier test de TSV.

3.3.4 Plots d'interconnexions

La conception de l'anneau d'interconnexions tout autour de la couche d'étouffement (figure 3.25) a été faite en étroite collaboration avec la couche de traitement numérique de signaux. En effet, chacune des couches possède son propre anneau d'interconnexions, mais doit toutefois composer avec le fait que les deux anneaux sont collés l'un face à l'autre (en périphérie sur la 2^e et 3^e couche de la figure 3.1). Les connexions pour chacune des deux couches passent toutes par l'arrière de la couche d'étouffement, là où la couche de photodiodes ne se rend pas puisqu'elle est de plus petites dimensions que celle d'étouffement. Les soudures de fils viennent se connecter sur les plots de métal à l'arrière de la couche d'étouffement avant de descendre à l'intérieur de la couche par les TSV. Une connexion destinée à la couche de traitement numérique de signaux arrive donc tout d'abord sur

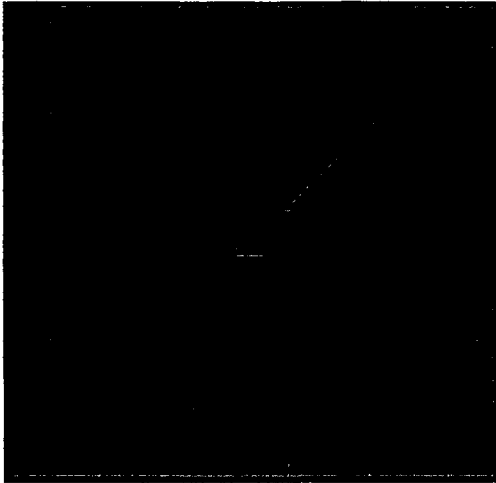


Figure 3.30 Dessin des masques du deuxième test de TSV.

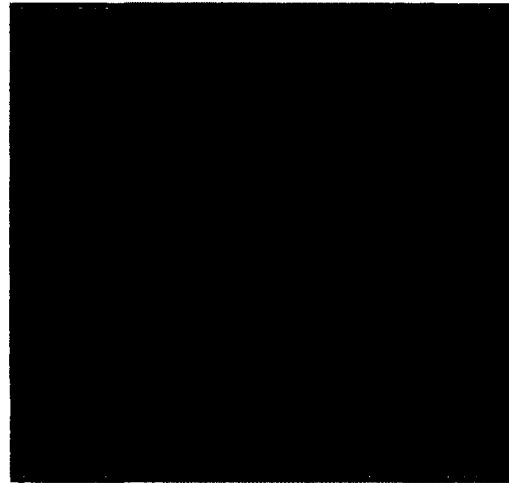


Figure 3.31 Dessin des masques du troisième test de TSV.

un plot de la couche d'étouffement et se transfère sur le plot de la couche de traitement numérique de signaux collé directement face à lui. Pour une connexion devant se rendre sur la couche d'étouffement seulement, le signal s'arrête sur un plot de cette couche et aucun plot ne se retrouve face à lui sur la couche de traitement numérique de signaux. C'est pourquoi la conception de cet anneau a été faite de paire pour les deux couches, afin de s'assurer que les plots soient placés exactement au même endroit sur les deux couches, l'une en miroir à l'autre. Pour éviter qu'une erreur puisse se glisser dans cette partie de la conception, les mêmes éléments ont été utilisés. Effectivement, la démarche employée se sert d'un fichier texte dans lequel est indiqué la position de chacun des plots, ainsi que le signal arrivant sur chacun de ces plots, sur la couche d'étouffement ou sur la couche de traitement numérique de signaux. En utilisant le même fichier, on peut s'assurer que les deux anneaux seraient complémentaires. La même technique est également utilisée pour toutes les étiquettes de noms de signaux devant passer par les plots d'interconnexion sur la sixième couche de métal.

Ces fichiers sont ensuite compilés grâce au logiciel *Synopsys* qui synthétise le code VHDL en portes logiques. Le fichier résultant est à son tour compilé par le logiciel *Encounter* qui s'occupe de transformer l'information en dessins de masques, en faisant le placement physique de chacun des composants. Le tout peut finalement être importé à partir du logiciel *Cadence*, intégrant le tout au projet.

Le métal à l'arrière de la couche d'étouffement est utilisé pour des plots d'interconnexions autres que l'anneau mentionné précédemment. En effet, c'est par l'intermédiaire de ce mé-

tal que se fait le contact entre la couche de PAMP et les TSV de la couche d'étouffement. La distance centre à centre entre les circuits étant de 50 μm , il était important de laisser un certain espacement entre chaque métallisation. L'ouverture de la passivation était également soumise à des règles de conception par le procédé de fabrication. La métallisation a par conséquent été faite avec des octogones de 40 μm de diamètre qui ont été répétés uniformément pour couvrir la majeure partie de l'arrière de la couche (figure 3.32). Cependant, la passivation n'a été ouverte qu'aux points d'intérêts afin de faire les connexions. C'est également par ces plots de métal que les drains des DEMOS sont accessibles, tout comme les entrées et sorties de tests pour les circuits d'étouffement entre plots et les structures de tests des TSV. La figure 3.33 présente tous les points d'entrées et de sorties des structures de tests pour la couche d'étouffement en plus des différents points de connexions pour l'alimentation. Les chiffres en rouge dans le coin supérieur gauche de chacune des matrices donnent leur numéro respectif.

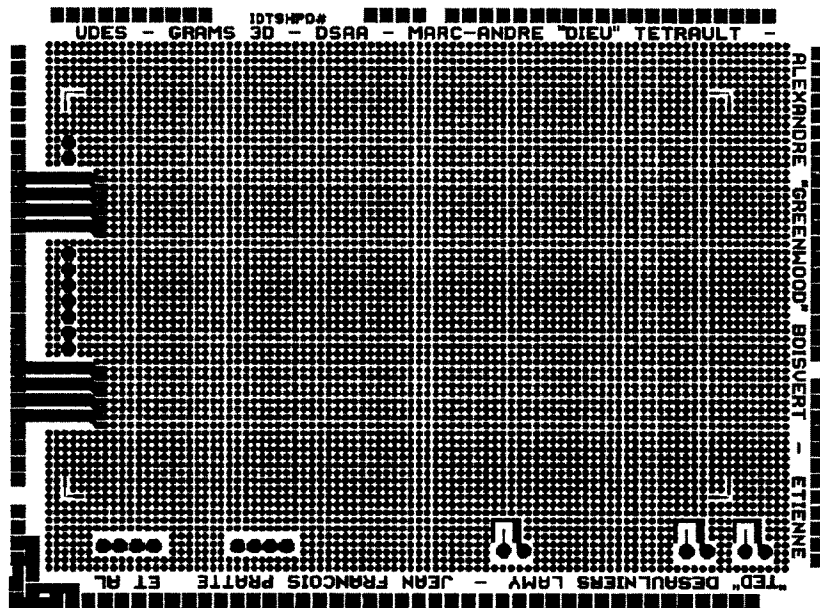


Figure 3.32 Dessin des masques de la métallisation derrière la couche d'étouffement.

3.3.5 Marques d'alignements et remplissage de métal

Les six matrices de circuits, les structures de test, ainsi que l'anneau d'interconnexions étant tous placés, la dernière étape consiste en la finition de la couche. Tout d'abord, afin d'aider le placement de la matrice de PAMP derrière la couche d'étouffement, des marques d'alignement sont placées dans quatre coins sur le métal de derrière. Ces marques ont été

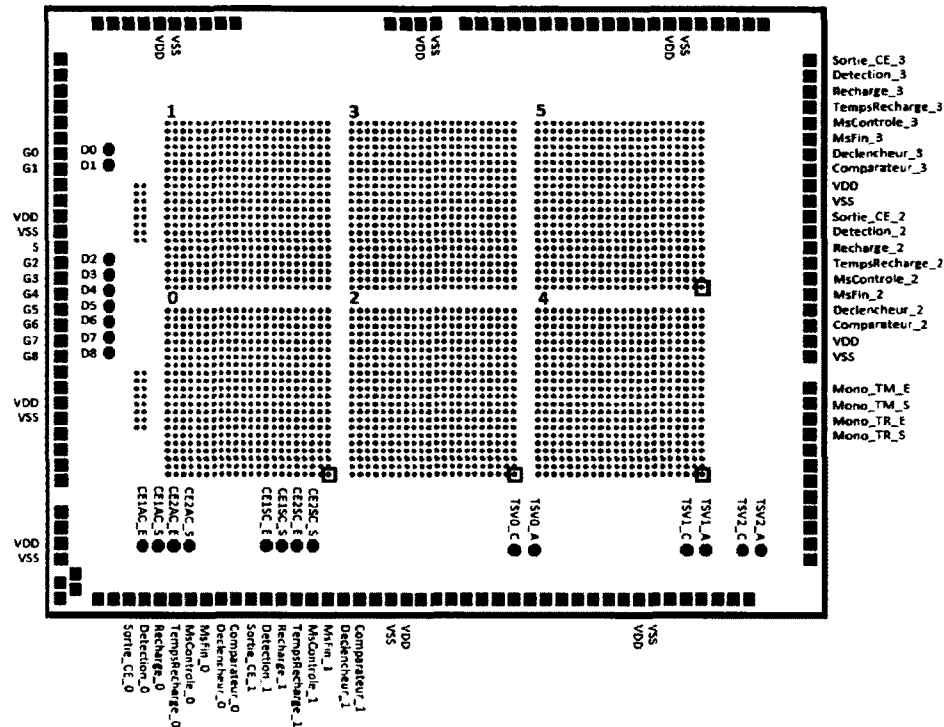


Figure 3.33 Dessin des masques de l'ouverture de la passivation pour les plots d'interconnexions.

faites en inverse de celles se retrouvant sur la couche des photodiodes, de façon à former une croix précise dans chacun des coins une fois les deux couches alignées. Tel que montré à la figure 3.34, la marque est faite de trois grosseurs différentes de métaux rectangulaires. Le plus petit rectangle présente une longueur de 20 μm et une largeur de 4 μm , le second double les dimensions avec une longueur de 40 μm et une largeur de 8 μm , alors que le troisième triple les dimensions pour atteindre 60 μm de long par 16 μm de large.

Finalement, à l'intérieur de la couche d'étouffement, un maillage de tous les métaux doit être fait, afin de bien propager l'alimentation. De plus, le procédé de fabrication de circuits intégrés impose une certaine densité de tous les matériaux, par des limites minimales et maximales de pourcentage d'aire occupée. Afin de faciliter le remplissage de l'espace à combler sur la couche et de respecter ces normes, une cellule de $50 \times 50 \mu\text{m}^2$ a été créée. Cette cellule fait un quadrillage de tous les métaux en utilisant une densité quelque peu sous la limite maximale du pourcentage d'aire occupée. Une répétition de cette cellule a permis rapidement de combler l'espace inutilisé, tout en assurant une densité uniforme de tous les matériaux sur la couche.

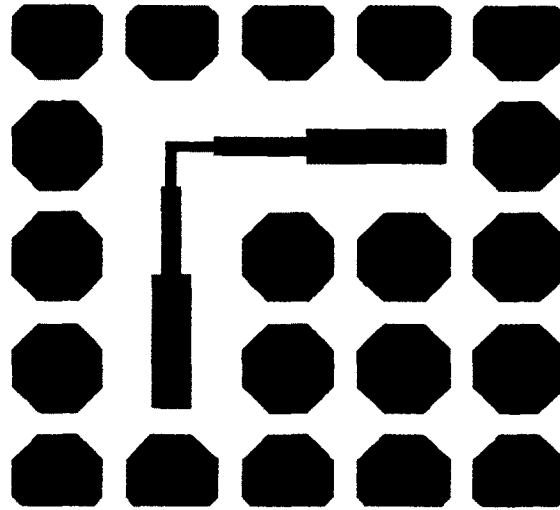


Figure 3.34 Marques d'alignement.

3.4 PCB et plan de tests

Afin de pouvoir communiquer avec la puce et de faire les tests requis, un circuit imprimé (*Printed Circuit Board*, PCB) est nécessaire pour agir en tant qu'intermédiaire. Pour permettre une réutilisation des ressources dans des projets futurs, deux cartes ont été créées, soit une carte mère et une carte fille, montrées respectivement aux figures 3.35 et 3.36. La carte mère possède les plus gros composants et les plus coûteux pouvant être réutilisés, tels l'alimentation, le FPGA, la communication vers un ordinateur, etc. Sur la carte fille repose les modules requis pour abaisser la tension de l'alimentation à celles nécessaires pour la puce développée, en plus des parties utiles pour les tests sans oublier, bien évidemment, la puce en elle-même. Ainsi, la carte mère pourra être réutilisée d'un projet à l'autre, en ne changeant que la carte fille pour s'adapter au projet en cours.

Les structures de tests planifiées permettent la vérification de différents aspects du circuit. Cette vérification devra être faite à la réception des puces, puisqu'elles sont toujours en fabrication en date d'écriture de ce mémoire. Ce plan de tests présente six structures différentes : la fonctionnalité des circuits d'étouffement, le courant de la cathode d'une PAMP, la gigue des circuits d'étouffement, la mesure des temps des monostables, la caractérisation des transistors DEMOS et la tension de claquage des TSV. Afin de bien comprendre où sont situés les différents circuits de tests, veuillez vous référer à la figure 3.33 pour les différents plots d'entrées et sorties.

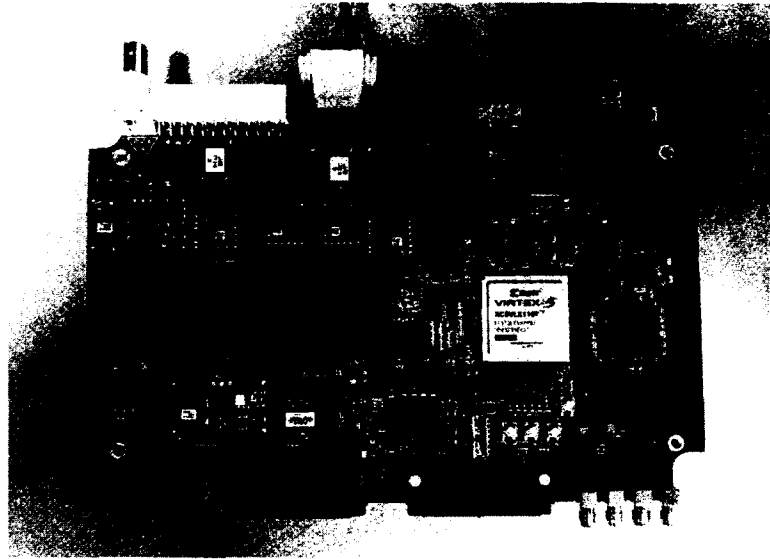


Figure 3.35 Photo de la carte mère.

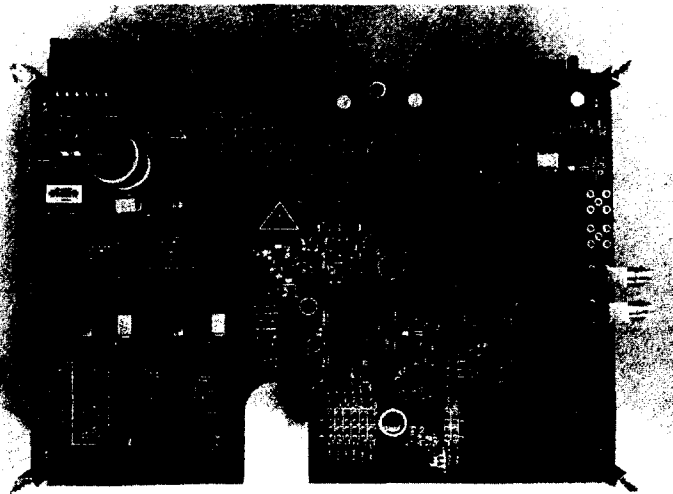


Figure 3.36 Photo de la carte fille.

3.4.1 Fonctionnalité des circuits d'étouffement

La vérification de la fonctionnalité des circuits d'étouffement se complète à travers structures différentes de test. Tout d'abord, quatre circuits ont été conçus pour vérifier huit nœuds critiques aux circuits d'étouffement. Les signaux observés sont : *Compareur*, *De-clencheur*, *MsFin*, *MsControle*, *TempsRecharge*, *Recharge*, *Detection*, *Sortie_CE*. Ces circuits sont encadrés en rouge sur la figure 3.33. Les circuits 0 et 1, intégrés aux matrices 0 et

2 respectivement, sont des circuits d'étouffement avec des transistors 3,3 V en entrée, alors que les circuits 2 et 3, intégrés aux matrices 4 et 5, ont des transistors DEMOS en entrée. Le contrôle pour ce test se fait par l'interface logicielle (figure 3.37), par l'intermédiaire du FPGA et de la couche de traitement numérique de signaux.

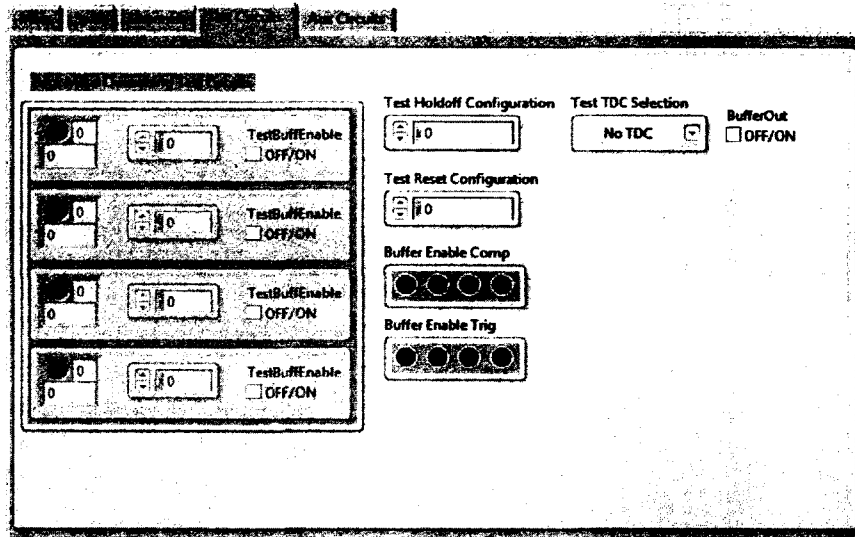


Figure 3.37 Interface logicielle pour le contrôle des circuits de test.

Tel qu'expliqué à la section 3.3.3, les deux premiers signaux sont amplifiés par une série d'inverseurs avant de sortir sur des plots d'interconnexions analogiques directement sur la couche d'étouffement. Les autres signaux passent quant à eux par la couche de traitement numérique de signaux et sortent sur des plots d'interconnexions numériques. Tous les signaux sont ensuite envoyés vers le FPGA se trouvant sur la carte mère. Les signaux *Compareur* et *Sortie_CE* étant les deux plus importants du circuit d'étouffement, ils possèdent également un point de lecture directement sur la carte fille, tels que montrés sur la figure 3.38. Le signal *Compareur* permet de s'assurer que la tension monte à l'entrée du comparateur de tension, alors que le signal *Sortie_CE* vérifie que le comparateur a bien détecté l'événement et que le circuit a répondu au signal, la logique de contrôle donnant le signal de sortie et le monostable du temps mort donnant la durée de ce signal. Ces signaux peuvent être observés avec l'oscilloscope LeCroy SDA 6000A disponible au GRAMS. Ces signaux suivent l'allure des courbes présentées à la figure 3.4, à l'exception que le signal *Compareur* observé à l'oscilloscope aura une forme d'onde carrée, en raison de son passage à travers la série d'inverseurs. Advenant un problème avec un circuit d'étouffement, les six autres signaux peuvent être observés grâce au FPGA et à l'interface logicielle, programmée à l'aide du logiciel *LabVIEW* de *National Instruments*.

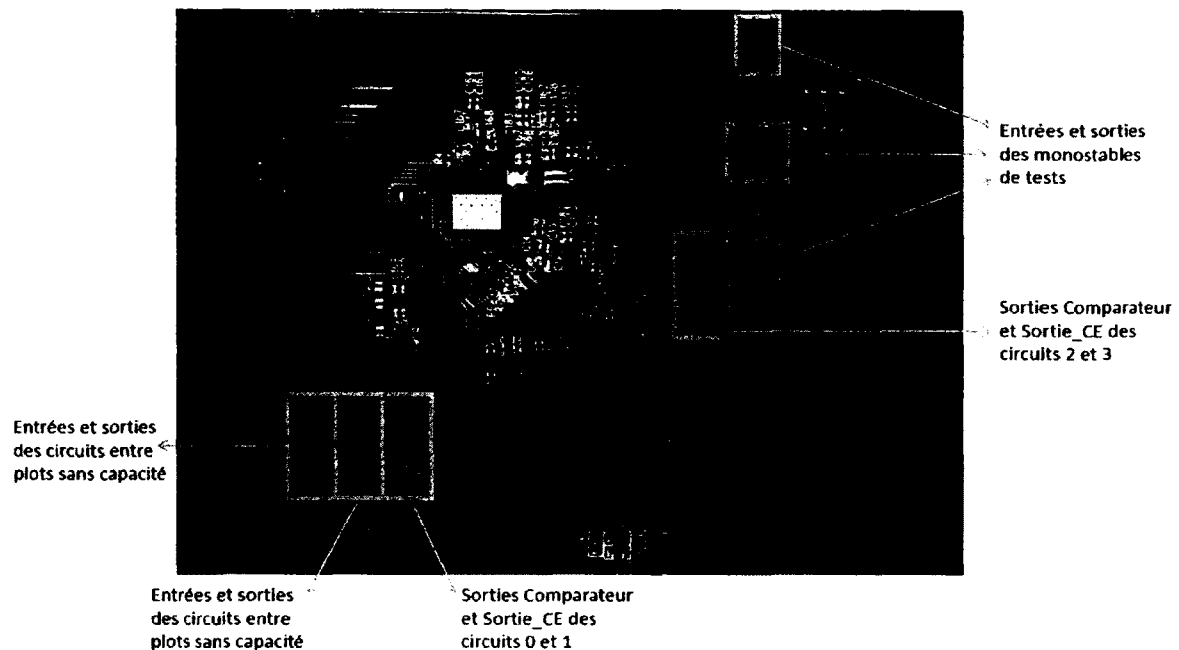


Figure 3.38 Agrandissement de la figure 3.36 montrant les points de lectures des différents tests sur la carte fille.

Tel que mentionné précédemment, la fonctionnalité des circuits d'étouffement peut être vérifiée de diverses façons. La première consiste à utiliser le générateur d'impulsions se retrouvant à l'intérieur du circuit d'étouffement ayant été présenté à la section 3.2.11. Avant de commencer la vérification, il faut s'assurer que le signal *Activateur* est au niveau haut et qu'une valeur a été donnée aux bits de configurations des monostables. Il faut également que le signal *GIDeclencheur* soit désactivé par un niveau bas. Le signal *GIR recharge* doit être activé pendant au moins un cycle d'horloge avant d'être redésactivé, afin de permettre la charge de la capacité MOS MN1 (figure 3.20). Le circuit doit ensuite être déclenché une première fois, afin de s'assurer qu'aucun nœud ne soit dans un état métastable et que le circuit se trouve dans les conditions désirées. Pour utiliser le générateur d'impulsions, il faut activer le signal *GIDeclencheur* pendant un cycle d'horloge, alors que le signal *GIR recharge* est désactivé. Une fois ce cycle terminé, le signal *GIDeclencheur* est désactivé, alors que le signal *GIR recharge* est à son tour activé pour un cycle d'horloge. Il est également important de vérifier que le signal de sortie *Sortie_CE* demeure désactivé à un niveau bas lorsque le signal *Activateur* est désactivé, et ce, même si le circuit est déclenché par le générateur d'impulsions.

La seconde façon de vérifier la fonctionnalité des circuits d'étouffement se fait grâce aux circuits entre plots. En utilisant un générateur de signaux, il est possible de vérifier que les

transistors d'entrée des circuits, soient MN0 et MN1 (figure 3.6), fonctionnent correctement et que le circuit répond au changement survenu à l'entrée à l'aide de l'oscilloscope branché à la sortie. Ce test permet de compléter la première méthode avec le générateur d'impulsions qui ne considère pas les deux transistors d'entrée. En connectant une source de tension à l'entrée du circuit d'éteuffement entre plots sans capacité, la tension de seuil nécessaire pour le déclenchement du circuit peut être vérifiée en faisant varier la tension d'entrée très lentement jusqu'à ce que la sortie change d'état.

La fonctionnalité des circuits d'éteuffement peut se vérifier d'une troisième façon, c'est-à-dire en utilisant les matrices de circuits avec la couche de PAMP branchée au-dessus. En mettant le montage dans une boîte noire, il est possible de vérifier que le circuit fonctionne une fois connecté à l'anode d'une photodiode simplement avec les déclenchements issus du bruit d'obscurité.

3.4.2 Gigue des circuits d'éteuffement

Plusieurs structures ont été prévues pour les mesures de giges des circuits d'éteuffement. Initialement, la structure choisie consistait à envoyer une impulsion à l'entrée des circuits entre plots de la figure 3.33, d'observer ce signal à l'oscilloscope avec une sonde active et de mesurer la différence en temps avec le signal de sortie. En répétant cet exercice et en accumulant les résultats, il est possible d'estimer la gigue du circuit grâce à l'histogramme de la variation des signaux de sortie entre eux. Lors de la conception des cartes PCB, après l'envoi des puces électroniques en fabrication, une puce de convertisseur temps-numérique (*Time-to-Digital Converter*, TDC) ACAM a été ajoutée sur la carte mère [Precision Measurement Technologies, 2013]. Le FPGA peut donc envoyer un signal à l'entrée du circuit entre plots et démarrer le TDC en même temps avec le même signal. La sortie du circuit est envoyée au TDC afin d'arrêter sa mesure en temps.

Cependant, un élément n'a pas été pris en considération lors de la conception de cette structure de test. Afin que la sortie du circuit d'éteuffement puisse être lisible avec une sonde, le signal circule à travers une série d'inverseurs. Malheureusement, cette série d'inverseurs n'a pas été isolée entre deux plots d'interconnexion pour en faire la caractérisation. La gigue de ce tampon vient par conséquent s'ajouter à celle du circuit d'éteuffement et ainsi fausser la valeur lue, en ne sachant pas quelle portion de la valeur provient du circuit et laquelle provient de la série d'inverseurs. L'approche considérée pour résoudre ce problème au mieux consiste à mesurer la gigue du circuit d'éteuffement avec la série d'inverseurs. Ensuite, en simulant la gigue du circuit d'éteuffement et celle du tampon grâce aux extractions avec parasites du logiciel *Cadence*, il devrait être possible d'approximer la part

due au circuit d'étouffement et celle due au tampon. Afin d'avoir un deuxième regard sur la valeur trouvée, le même principe peut être appliqué à un circuit de test des matrices. Cela requiert toutefois une photolithographie en salle blanche afin d'ouvrir plus grand le plot d'interconnexion qui est plus petit à cet endroit que ceux des circuits entre plots. Une fois cette étape faite, la même impulsion peut être envoyée à l'entrée du circuit qui cette fois envoie son signal de sortie sur un plot d'interconnexion numérique, par un des quatre signaux *Sortie_CE_x* (figure 3.33). La même approche est utilisée, soit de simuler grâce aux extractions avec parasites la gigue du plot d'interconnexion numérique, afin de pouvoir approximer sa part dans la valeur mesurée.

3.4.3 Courant de la cathode d'une PAMP

Une autre structure de tests a été prévue pour les circuits d'étouffement de tests des matrices 2 et 4 (figure 3.33). Cette structure consiste à faire une lecture du courant circulant par la cathode d'une PAMP via l'alimentation de la haute tension pour les matrices de PAMP à partir de la carte fille. En effet, pour l'alimentation des matrices 2 et 4 de PAMP, la figure 3.39 montre le schéma électrique correspondant au chemin parcouru sur la carte fille avant d'atteindre la couche de PAMP. La capacité C_1 permet de stabiliser la haute tension d'alimentation provenant de la source. L_1 correspond à une inductance placée pour empêcher une variation rapide de courant vers la source, isolant ainsi les capacités C_2 , C_3 , C_4 et la matrice de PAMP d'avec le reste du circuit. La matrice de PAMP se connecte aux capacités par le plot d'interconnexion grâce à un fil (*wirebond*).

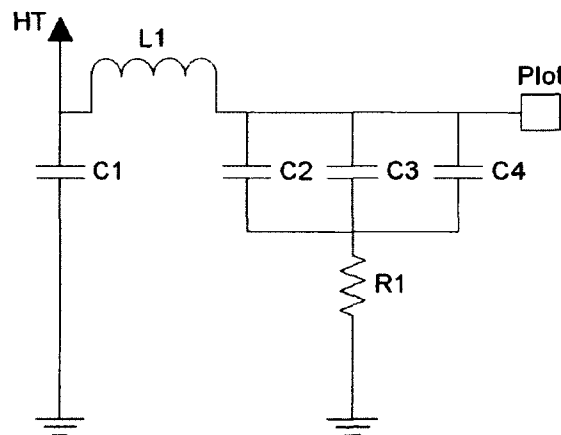


Figure 3.39 Schéma électrique du point de test du courant de la cathode d'une PAMP.

L'objectif de cette structure est de tester seulement le circuit d'étouffement de test de la matrice choisie. Comme toutes les PAMP de la matrice sont reliées en parallèle à la même alimentation, tous les circuits d'étouffement doivent être désactivés, à l'exception du circuit de test, afin de faire la mesure désirée. Pour ce faire, le signal *Activeur* des circuits d'étouffement doit présenter un niveau bas. Ainsi, une fois que toutes les PAMP se sont déclenchées et que les circuits les ont étouffées, les circuits d'étouffement les gardent dans cet état jusqu'à ce que l'utilisateur décide de les réactiver. Il ne reste par conséquent qu'une PAMP et son circuit d'étouffement d'actifs. Lorsque la PAMP sera déclenchée, l'inductance empêchant la variation rapide de courant, les capacités C_2 , C_3 et C_4 devront fournir les charges nécessaires à la PAMP. Cela provoquera un courant à travers la résistance R_1 de $100\ \Omega$ permettant de voir l'allure de la courbe du courant tiré par la cathode, et ce, sans devoir mesurer directement sur la haute tension. Cette courbe est observée en mesurant la tension aux bornes de la résistance R_1 grâce à un oscilloscope branché aux connecteurs de la figure 3.40. Afin de bien bloquer la variation de courant vers la source, l'inductance a été choisie pour présenter une réactance approximativement dix fois plus grande que les résistances équivalentes séries (*Equivalent Series Resistance*, ESR) des capacités mises en parallèle, ces dernières en série avec la résistance R_1 . Cette structure de test permet d'évaluer le temps d'étouffement, ainsi que le temps de recharge de la PAMP. Une fois ce test terminé, afin de pouvoir bien utiliser la matrice de PAMP, l'inductance L_1 et la résistance R_1 sont remplacées par deux résistances de $0\ \Omega$.

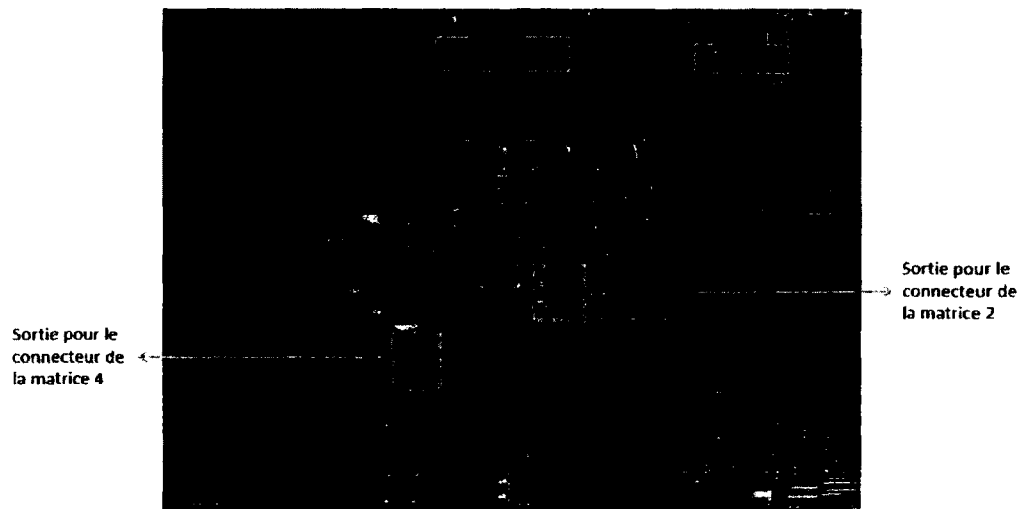


Figure 3.40 Points de lectures des tests du courant de la cathode d'une PAMP.

3.4.4 Mesure des temps des monostables

Tel que décrit à la section 3.3.3, un monostable de temps mort et un monostable de temps de recharge ont été insérés sur la couche d'étouffement pour des tests. Cette structure permettra de vérifier si les temps obtenus des différentes combinaisons concordent avec ceux simulés, présentés aux tableaux 3.1 et 3.2. Le FPGA, par le biais de l'interface logicielle, envoie un niveau haut aux entrées *Mono_TM_E* et *Mono_TR_E* (figure 3.33), démarrant ainsi les monostables. Ces derniers retournent un signal haut par les plots *Mono_TM_S* et *Mono_TR_S*. Des connecteurs ont été prévus sur la carte fille près de la puce (figure 3.38), afin de permettre la lecture de ces différents signaux par oscilloscope. Les 16 possibilités pour chacun des monostables sont choisies à partir de l'interface logicielle, permettant de toutes les mesurer l'une à la suite de l'autre.

3.4.5 Caractérisation des transistors DEMOS

Les transistors DEMOS fabriqués selon la méthode présentée à la section 2.3, 9 structures différentes, décrites à la section 3.3.3, ont été placées pour en faire la caractérisation. Les divers tests permettant cette caractérisation se font par l'intermédiaire d'une station sous pointes, disponible au CRN2 de l'Université de Sherbrooke. Les quatre pointes de la station font un contact direct avec les plots métalliques à l'arrière de la couche d'étouffement (figure 3.33). Chaque transistor DEMOS possède sa propre connexion à son drain et à sa grille, soient D0 à D8 et G0 à G8. Par contre, toutes les sources sont connectées à un seul et même plot, c'est-à-dire S, limitant ainsi le nombre d'interconnexions utilisées. La quatrième et dernière pointe se connecte au plot amenant la masse à la couche d'étouffement.

Le premier test à effectuer correspond à faire les différentes courbes I_d/V_{ds} selon un balayage de V_{gs} de 0 à 3,3 V en utilisant l'analyseur paramétrique Keithley 4200. Ensuite, à l'aide de deux générateurs de signaux, c'est-à-dire l'un branché à la grille et le second au drain, un test de stress peut être fait. Ce test consiste à reproduire les conditions d'utilisation des transistors à l'entrée des circuits d'étouffement. Dans le cas du transistor de détection/étouffement, le signal au drain varie comme le nœud *TSV_Tiers2* et celui de sa grille comme le nœud *Detection* (figure 3.4). Dans le cas du transistor de recharge, le même signal que le transistor précédent varie à son drain, alors que le signal *Recharge* contrôle sa grille. Les deux générateurs de signaux font osciller les transistors l'un à la suite de l'autre pendant quelques jours, afin d'observer l'impact à plus long terme de l'utilisation de ces transistors dans ces conditions. Finalement, le dernier test consiste à augmenter la tension

au drain des transistors pour mesurer leur tension de claquage. Ce test doit cependant être fait à la fin puisqu'il s'agit d'un test destructif.

3.4.6 Tension de claquage des TSV

Trois structures différentes de tests ont été incluses afin de vérifier la tension pouvant être atteinte avant un claquage en utilisant des TSV (section 3.3.3). En effet, dans l'optique d'une future version de la puce où la haute tension pour l'alimentation des PAMP proviendrait de l'arrière de la puce et devrait passer à travers la couche de traitement numérique de signaux, de la couche d'étouffement et de la couche de PAMP par les TSV, il était dans l'intérêt du GRAMS de connaître la tension pouvant être tolérée par les TSV et les couches de métallisation du procédé de fabrication. En utilisant l'analyseur paramétrique Keithley 4200, qui peut atteindre une tension maximale de 200 V, et la station sous pointes, les contacts avec les structures de tests se font directement par la métallisation à l'arrière de la couche d'étouffement par les plots de la figure 3.33. Le plot *TSVx_A*, où x correspond au numéro de la structure, amène la masse sur un anneau autour de l'autre TSV qui est branché à la tension d'alimentation. La tension est alors augmentée sur le TSV central par le plot *TSVx_C* jusqu'à obtenir un claquage entre les deux. Le même principe est utilisé pour les trois structures.

CHAPITRE 4

RÉSULTATS ET DISCUSSIONS

En raison de retards de production par les fabricants, les circuits physiques ne sont toujours pas disponibles en date d'écriture de ce mémoire, soit deux ans après la soumission pour fabrication. C'est pourquoi les résultats et l'atteinte des objectifs seront basés sur les simulations produites avec le logiciel *Cadence*.

Tout d'abord, le circuit doit atteindre les principaux buts d'un circuit d'étouffement, tels que vus précédemment à la section 2.2. Le comparateur de tension permet de détecter l'avalanche, alors que la logique de contrôle envoie un signal à la sortie synchrone avec le changement de courant. L'utilisation de transistors permet une charge variable pour étouffer rapidement de manière passive la photodiode. La branche de recharge permet une repolarisation active de la photodiode. Tous les buts visés ont par conséquent été atteints. Tout ceci ayant été développé à l'intérieur d'une architecture 3D, il est donc possible d'affirmer que l'objectif général a été atteint selon les simulations, soit de concevoir et intégrer un circuit d'étouffement pour une photodiode à avalanche en mode Geiger pour l'architecture 3D d'un nouveau scanner TEP.

Rappelons qu'il y a quatre objectifs spécifiques pour ce projet de maîtrise. Le premier consiste à concevoir le circuit d'étouffement, en utilisant le procédé de fabrication de Tezzaron/Chartered de CMOS 130 nm 3D, dans un espace de $50 \times 50 \mu\text{m}^2$, soit le même espace que pour une photodiode sur la couche au-dessus. Tels que montrés aux figures 3.22 et 3.24, les deux types de circuits entrent dans les limites imposées par l'objectif. Les circuits ont ensuite pu être copiés dans une matrice de 22×22 circuits, afin de reproduire la même matrice que celle de PAMP de la couche supérieure. Ces matrices couvrent ainsi tout l'espace sous la structure cristalline du scintillateur, soit $1,1 \times 1,1 \text{ mm}^2$, ne laissant que le dernier $50 \mu\text{m}$ en périphérie du scintillateur. Le premier objectif spécifique est donc comblé.

Le second objectif spécifique demande un contrôle du temps mort de la photodiode et également d'offrir différentes possibilités de durées de temps mort à partir de bits de configuration provenant de la couche de traitement numérique de signaux. Le monostable de temps mort vu à la section 3.2.6 permet effectivement au circuit de contrôler la durée pour laquelle la photodiode est gardée étouffée. De plus, grâce aux quatre bits de configuration,

16 temps morts différents sont disponibles, variant de 11,20 ns à 849,95 ns selon les simulations, tels que présentés au tableau 3.1. Le second objectif spécifique reçoit lui aussi une note positive quant à sa résolution.

Le troisième objectif spécifique stipule que le circuit doit tolérer la plus grande tension d'excès possible avec la technologie de Tezzaron/Chartered de CMOS 130 nm 3D. Pour arriver à cette fin, des transistors DEMOS ont été développés, tels que vus à la section 2.3. Le rendement de ces transistors ne peut cependant être commenté, puisqu'en l'absence des puces en date d'écriture, leur caractérisation n'a pu être faite. Il est par conséquent impossible de confirmer avec certitude que l'objectif a été atteint. Par contre, en se basant sur les résultats présentés dans l'article de référence de cette structure [Santos *et al.*, 2007], des tensions de claquages entre 23 V et 25 V ont été mesurées pour une alimentation de grille de 3,3 V, alors que leurs transistors étaient de procédé 350 nm. Cela permet donc raisonnablement de prévoir que la tension de claquage d'un transistor 3,3 V régulier sera dépassée avec cette architecture. Le troisième objectif devrait donc théoriquement être atteint.

Le quatrième et dernier objectif spécifique mentionne que le circuit doit avoir une faible consommation de puissance. Cette consommation est calculée grâce aux simulations de l'alimentation du circuit d'étouffement, en omettant les charges de l'avalanche. Cette omission se doit au fait que les charges proviennent de l'alimentation des photodiodes et dépend également des caractéristiques de la PAMP, en plus de la tension d'excès à laquelle la photodiode est opérée. Lors d'une simulation avec parasites, avec le temps mort et le temps de recharge les plus courts, un délai approximatif de 30 ns se déroule entre le moment auquel le photon est absorbé par la photodiode et le moment auquel la photodiode est prête à une nouvelle détection. Supposant un scénario où la photodiode se déclencherait à toutes les 30 ns pendant une seconde complète, cela supposerait un taux de comptage d'environ 33 Mcps. En intégrant les charges circulant par l'alimentation du circuit d'étouffement, il en résulte approximativement 3 pC pour une période de 30 ns, soit 100 pC par seconde. Avec la tension d'alimentation à 3,3 V, le circuit d'étouffement développé consomme en moyenne 330 pW pour ce taux de comptage. Par contre, cela ne représente pas une consommation réaliste pour une utilisation en TEP. Selon les connaissances acquises au GRAMS, il se produit approximativement entre 2000 et 3000 événements par seconde pour chaque scintillateur de $1,1 \times 1,1 \text{ mm}^2$. Pour chacun de ces événements, une PAMP reçoit en moyenne 2,5 photons. En supposant le pire cas de 3000 événements par seconde par scintillateur, le circuit d'étouffement serait par conséquent déclenché environ 7500 fois par seconde. Le circuit d'étouffement utilisant approximativement 3 pC par déclenchement, ce

dernier consomme en moyenne 74,25 nW pour un taux de comptage de 7500 kcps. Comme il y a 484 PAMP par matrice, cela donne donc une consommation de 35,9 μ W pour chaque matrice jumelée à un scintillateur. Un scanner TEP pour petits animaux contenant environ 37000 scintillateurs, le circuit d'étouffement contribue de 1,33 W sur la consommation totale du scanner. Ces résultats selon les simulations permettent d'atteindre l'objectif.

La question de recherche demande quant à elle que le circuit d'étouffement possède une faible gigue temporelle. Grâce au logiciel *Cadence*, la gigue temporelle due à l'électronique du circuit a pu être simulée. La simulation donne un résultat de 0,41 ps. Cependant, il faut également prendre en considération le bruit sur l'alimentation pouvant influencer cette gigue. Une variation de ± 100 mV a été mise sur l'alimentation VDD et sur la masse, afin d'observer la variation sur le signal de sortie. La gigue due à la variation sur l'alimentation VDD résulte à 2,70 ps, alors que la variation sur la masse donne une gigue de 6,95 ps. Ces différentes giges temporelles n'étant pas corrélées, elles doivent être additionnées en quadrature, donnant un résultat final de 7,47 ps pour la gigue temporelle du circuit. Avec ces éléments en main, on peut répondre à la question de recherche :

« Est-il possible d'étouffer une photodiode à avalanche en mode Geiger avec un circuit d'étouffement en technologie CMOS sous-micrométrique contraint en espace à la taille de la photodiode dans une architecture 3D, et ce, le plus rapidement possible, avec une gigue temporelle minimale, tout en contrôlant son temps mort et en limitant la consommation de puissance, pour un module de photodétection avec résolution monophotonique ? »

À la lumière des résultats obtenus par l'objectif général et les objectifs spécifiques, la question de recherche récolte une réponse affirmative en considérant les simulations faites avec le logiciel *Cadence*. Cependant, cette réponse devra être confirmée suite aux tests lors de la réception des puces.

CHAPITRE 5

CONCLUSION

Plusieurs domaines de l'industrie, plus particulièrement le domaine médical, utilisent la détection de lumière à des fins d'imagerie. Cette technologie photonique nécessite des détecteurs de plus en plus rapides, afin d'améliorer la qualité et la précision de l'image. Le GRAMS, dans une nouvelle version de scanner TEP, opte pour une architecture en trois dimensions et des photodiodes à avalanche opérées en mode Geiger pour améliorer cette facette de leur scanner. En mode Geiger, les photodiodes à avalanche ont besoin d'un circuit d'étouffement dans le but de détecter son avalanche, de l'arrêter et d'empêcher l'emballement de la photodiode, de lever un drapeau à l'électronique de logique pour signaler l'événement et de repolariser la photodiode.

Connaissant la mise en contexte du projet et les objectifs à atteindre, une revue de la littérature a été présentée, faisant tout d'abord un tour d'horizon des photodétecteurs : les tubes photomultiplicateurs, les photodiodes à avalanche en mode linéaire, les photodiodes à avalanche opérées en mode Geiger et les photomultiplicateurs sur silicium. Les PAMP présentent une faible gigue temporelle tout en étant compact, facile à matricer et insensible aux champs magnétiques, les rendant très attrayant pour un scanner TEP. Ensuite, après avoir expliqué les principes du circuit d'étouffement, différentes architectures de ces circuits ont été montrées : passif, actif, mixte, avec comparateur de courant et à charge variable. Le VLQC offre les avantages d'un étouffement passif et d'une recharge active, tout en étant très compact. Les transistors à drain étendu (DEMOS) ont été présentés par la suite, afin de tolérer une plus haute tension à l'entrée du circuit d'étouffement et ainsi améliorer la probabilité de détection de photons de la PAMP. La section 2.4 portait sur l'intégration 3D, en débutant avec l'historique amenant à l'intégration 3D, avant de se concentrer sur les interconnexions verticales. L'intégration 3D permet de réduire les délais d'interconnexions, tout en augmentant la densité d'électronique par surface. Également, cela permet d'avoir le circuit d'étouffement près de la PAMP sans trop diminuer le ratio de surface photosensible. La dernière section de l'état de l'art présentait des modèles de simulation pour les PAMP et les interconnexions verticales, afin de permettre une meilleure conception du circuit d'étouffement. Une fois cette revue de l'état de l'art faite, le chapitre 3 a présenté le circuit conçu à l'intérieur du projet. Le fonctionnement global du circuit a tout d'abord été expliqué avant de passer à une description plus détaillée de

chacun des modules : la modélisation des PAMP, la modélisation des TSV, l'électronique frontale avec des transistors de 3,3 V, l'électronique frontale avec des transistors DEMOS, la bascule, le monostable de temps mort, le monostable de temps de recharge, l'amorce et les références de courant, le traducteur de niveau, la logique de contrôle et le générateur d'impulsions. Les simulations d'une matrice et le schéma de la couche complète y ont également été expliqués. Une autre partie faisait quant à elle mention de la conception des masques. Cela englobait le dessin des masques du circuit d'étouffement en lui-même ainsi qu'en matrice, le dessin des masques des transistors DEMOS et des matrices à haute tension, les circuits de tests, les plots d'interconnexions, les marques d'alignement et le remplissage de métal. Les PCB de la carte mère et de la carte fille ont été introduits, avant de faire le tour du plan de tests. Finalement, les résultats de simulations ont été discutés dans le chapitre 4, puisqu'en n'ayant pas reçu les circuits fabriqués au moment d'écriture du mémoire, aucun test physique n'a pu être fait. La revue de tout le projet dans son ensemble ayant été faite, il reste, avec l'atteinte des objectifs posés pour ce projet, à répondre à la question de recherche afin de vérifier le succès de cette recherche.

L'objectif général de ce projet était de concevoir et d'intégrer un circuit d'étouffement pour une photodiode à avalanche opérée en mode Geiger pour l'architecture 3D d'un nouveau scanner TEP. Quatre objectifs spécifiques en découlaient ensuite. Le premier consistait de restreindre l'espace utilisé par le circuit à $50 \times 50 \mu\text{m}^2$ en utilisant le procédé de fabrication de Tezzaron/Chartered de CMOS 130 nm 3D. Dans l'optique de réduire le bruit postimpulsionnel, le second objectif spécifique visait à contrôler le temps mort de la photodiode et lui offrir plusieurs possibilités de temps mort grâce à des bits de configuration provenant de la couche de traitement numérique de signaux. Le troisième exigeait que le circuit puisse tolérer la plus grande tension d'excès possible avec la technologie CMOS 130 nm 3D de Tezzaron/Chartered. Le quatrième et dernier objectif spécifique demande que le circuit ait une faible consommation de puissance. Tel qu'il a été vu précédemment au chapitre 4, l'objectif général et le premier objectif spécifique ont été atteints. Toutefois, en raison d'un important retard dans la fabrication de la puce électronique développée, les derniers objectifs spécifiques n'ont pu être vérifiés physiquement par mesures. Cependant, en se basant sur les simulations et les résultats dans la littérature, les résultats devraient être positifs et les objectifs atteints. On peut donc répondre par l'affirmative à la question de recherche selon les simulations, mais devra être confirmée en procédant aux tests des structures une fois la puce reçue.

Toutefois, malgré le fait que le projet obtienne une note positive, plusieurs améliorations peuvent encore être apportées au circuit. Tout d'abord, beaucoup d'espace peut être sauvé

sur différents aspects. Le premier implique le monostable du temps de recharge. Au moment de la conception du circuit, les photodiodes à utiliser n'étaient pas encore caractérisées. Ne sachant pas si les PAMP réagiraient comme dans leurs simulations, il était important d'offrir différents temps de recharge s'étalant de 10 ns à 108 ns afin de s'assurer que la photodiode serait totalement rechargée avant de réactiver la détection. Cependant, comme les PAMP ont été caractérisées avec un autre circuit entre temps, le circuit d'étouffement pourrait utiliser un temps de recharge avec plage restreinte, voire fixe. Cela permettrait d'enlever les options de ce monostable et de récupérer de l'espace. Dans le même ordre d'idées, avec cette caractérisation des PAMP réalisée, le monostable de temps mort n'aura plus besoin d'une aussi grande flexibilité. Les capacités utilisées pourront être réduites et peut-être même utiliser moins de branches de drainage des charges. Suite à une révision du circuit après l'envoi en fabrication, il a été constaté que les références de courant seraient plus stables en se branchant directement au miroir de courant du circuit de l'amorce par le transistor M5 (figure 3.12) plutôt que par le transistor M6, comme cela a été fait dans la présente architecture. Outre les monostables, une autre façon de réduire l'espace utilisé par le circuit serait d'alimenter le circuit avec une tension de 1,5 V plutôt que 3,3 V. Ainsi, les transistors 1,5 V pourraient être utilisés au profit de leur taille réduite comparativement aux transistors 3,3 V. Également, en ayant la même tension d'alimentation que la couche de traitement numérique de signaux, les sept traducteurs de niveau pour les bits de configuration perdraient leur raison d'être, libérant encore une fois de l'espace. Un autre élément pouvant être amélioré sur le circuit se situe au niveau de la détection de l'avalanche. La tension à l'entrée du comparateur de tension doit monter à quelques centaines de millivolts afin de provoquer le déclenchement du circuit. En développant un circuit de détection avec un seuil plus bas, l'action d'étouffement pourrait être plus rapide que dans le cas présent. Un détecteur avec un seuil ajustable pourrait peut-être même être utile pour certaines applications.

Plusieurs applications pourraient bénéficier de cette technologie pour la détection photonique. En effet, la technologie utilisée dans le présent projet permet la détection des premiers photons arrivés au détecteur, en plus d'avoir un maximum de surface sensible étant donné que son électronique de contrôle se situe à l'arrière du détecteur. Cependant, la dissipation de la chaleur devient problématique en intégration 3D. Les couches étant collées, la chaleur se retrouve à présent confinée entre les couches. La communauté scientifique devra donc se tourner vers une nouvelle problématique, soit le refroidissement des couches. La solution résiderait-elle dans les canaux microfluidiques ? Ce qui est certain, c'est que cette technologie possède un potentiel irréfutable et que la communauté scientifique devra approfondir les multiples possibilités qu'elle lui offre.

ANNEXE A

DIMENSIONS DES TRANSISTORS

Cette annexe présente les dimensions de tous les transistors formant un circuit d'étouffement tel que vu dans ce document. Ces informations étant confidentielles, veuillez contacter le Groupe de Recherche en Appareillage Médical de Sherbrooke (GRAMS) pour de plus amples renseignements sur cette annexe.

ANNEXE B

VALEURS CONFIDENTIELLES

Cette annexe présente les capacités mesurées pour un SPAD fabriqué par Benoit-Louis Bérubé du GRAMS dans un procédé de fabrication de Dalsa. Ces valeurs sont utiles pour les simulations de la section 3.2.1. Cette annexe contient également les variables confidentielles nécessaires pour les calculs des TSV présentés à la section 3.2.2. Les résultats des calculs de ces TSV se retrouvent également dans cette annexe. Finalement, les capacités offertes pour les MIMCAP et les MOSCAP du procédé de fabrication de Tezzaron/Chartered CMOS 130 nm 3D y sont présentées. Ces informations étant confidentielles, veuillez contacter le Groupe de Recherche en Appareillage Médical de Sherbrooke (GRAMS) pour de plus amples renseignements sur cette annexe.

ANNEXE C

DIAGRAMMES D'ÉTATS

Cette annexe présente les diagrammes d'états des quatre portes de la logique de contrôle.

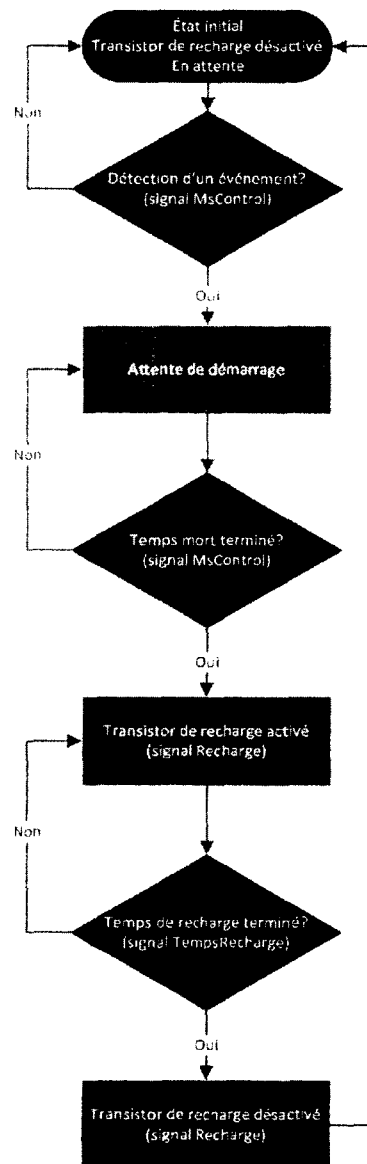


Figure C.1 Diagramme d'états du contrôle de recharge de la PAMP.

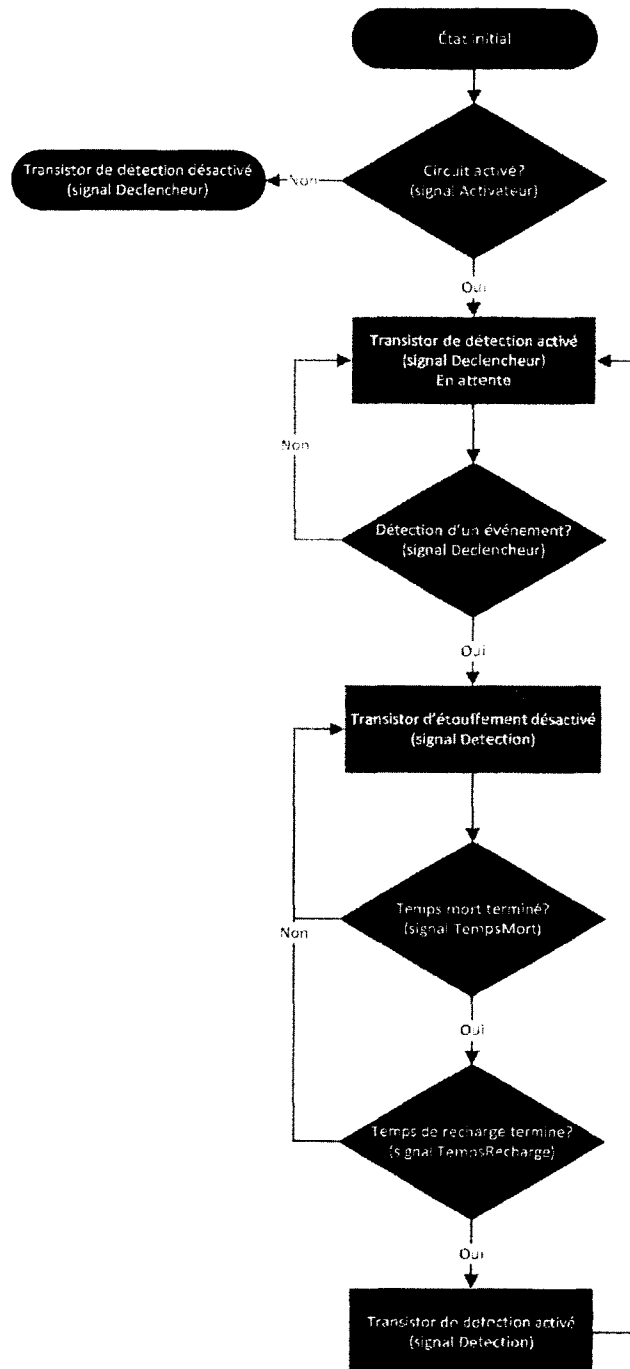


Figure C.2 Diagramme d'états du contrôle de détection/étouffement.

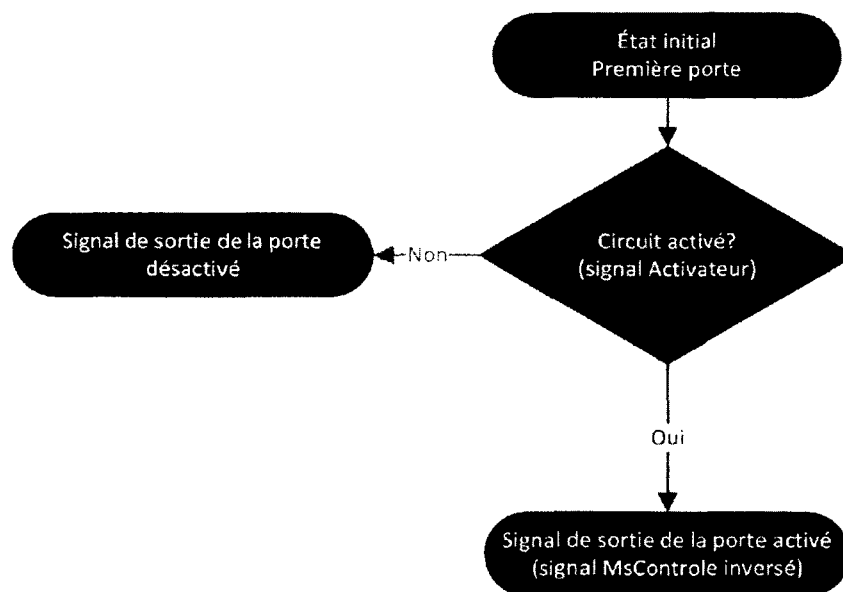


Figure C.3 Diagramme d'états du contrôle de sortie par l'activateur.

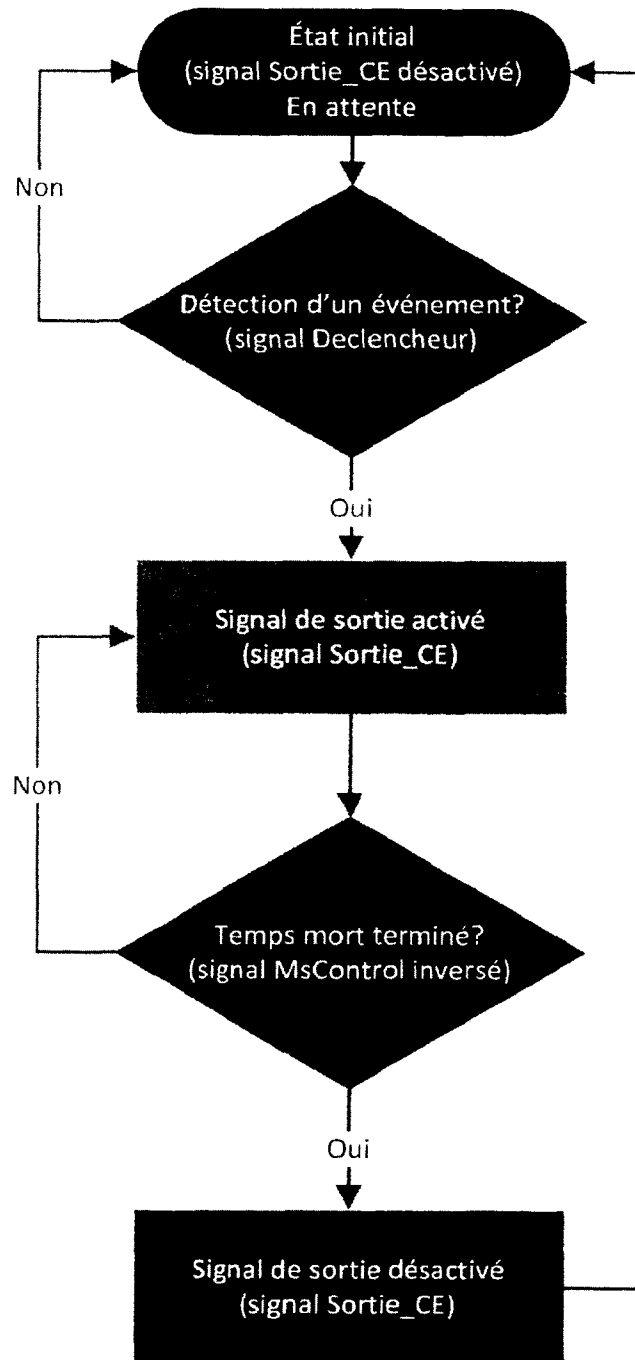


Figure C.4 Diagramme d'états du contrôle de la sortie du circuit d'étouffement.

LISTE DES RÉFÉRENCES

- Aarts, A., D'Halleweyn, N. et van Langevelde, R. (2005). A surface-potential-based high-voltage compact LDMOS transistor model. *IEEE Transactions on Electron Devices*, volume 52, numéro 5, p. 999–1007.
- Al-Ruwaihi, K. M. et Noras, J. M. (1995). Programmable CMOS current comparator circuit for analogue VLSI neural networks utilizing identical small-dimension MOS transistors. *International Journal of Electronics*, volume 78, numéro 2, p. 347–358.
- Allen, P. E. et Holberg, D. R. (2002). *CMOS Analog Circuit Design*, 2^e édition. Oxford University Press, 784 p.
- Aull, B., Burns, J., Chen, S., Felton, B., Hanson, H., Keast, C., Knecht, J., Loomis, A., Renzl, M., Soares, A., Suntharalingam, V., Warner, K., Wolfson, D., Yost, D. et Young, D. (2006). Laser radar imager based on 3D integration of geiger-mode avalanche photodiodes with two SOI timing circuit layers. Dans *2006 IEEE International Solid-State Circuits Conference, ISSCC, February 6, 2006 - February 9, MIT Lincoln Laboratory, Lexington, MA. Institute of Electrical and Electronics Engineers Inc, San Francisco, CA, United states*, p. 287–304.
- AutoNOMOS Labs (2011). *AutoNOMOS Labs*. <http://www.autonomos-labs.de/> (page consultée le 3 mai 2012).
- Bianchi, R. A., Monsieur, F., Blanchet, F., Raynaud, C. et Noblanc, O. (2008). High voltage devices integration into advanced CMOS technologies. Dans *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*. p. 1–4.
- Boffety, M. (2010). *Étude quantitative de la tomographie optique diffuse de luminescence. Application à la localisation de sources en imagerie moléculaire*. Thèse de doctorat, École Centrale Paris, Châtenay-Malabry, Île-de-France, France, 136 p.
- Charbon, E. (2008). Towards large scale CMOS single-photon detector arrays for lab-on-chip applications. *Journal of Physics D : Applied Physics*, volume 41, numéro 9, p. 094010 (9 pp.).
- Cherry, S. R., Sorenson, J. A. et Phelps, M. E. (2003). *Physics in Nuclear Medicine*, 3^e édition. Saunders, 544 p.
- Cova, S., Ghioni, M., Lacaita, A., Samori, C. et Zappa, F. (1996). Avalanche photodiodes and quenching circuits for single-photon detection. *Applied Optics*, volume 35, numéro 12, p. 1956–1976.
- Cova, S., Ghioni, M., Lotito, A., Rech, I. et Zappa, F. (2004). Evolution and prospects for single-photon avalanche diodes and quenching circuits. *Journal of Modern Optics*, volume 51, numéro 9-10, p. 1267–1288.

- Dalla Mora, A., Tosi, A., Tisa, S. et Zappa, F. (2007). Single-photon avalanche diode model for circuit simulations. *IEEE Photonics Technology Letters*, volume 19, numéro 23, p. 1922–1924.
- Dubois, A. (2007). Imagerie optique des milieux biologiques. Dans Institut d'Optique Graduate School, *IOGS - Catalogue des Cours*. <http://paristech.iota.u-psud.fr/site.php?id=209&fileid=486> (page consultée le 23 mai 2012).
- Eraerds, P., Legre, M., Rochas, A., Zbinden, H. et Gisin, N. (2007). SiPM for fast photon-counting and multiphoton detection. *Optics Express*, volume 15, numéro 22, p. 14539–14549.
- Finkelstein, H., Hsu, M., Zlatanovic, S. et Esener, S. (2007). Performance trade-offs in single-photon avalanche diode miniaturization. *Review of Scientific Instruments*, volume 78, numéro 10.
- Fontaine, R. (2010). Chapter 6. Design Considerations for Positron Emission Tomography (PET) Scanners Dedicated to Small-Animal Imaging. Dans Iniewski, K., *Electronics for Radiation Detection*. CRC Press, p. 151–177.
- Fontaine, R., Belanger, F., Viscogliosi, N., Semmaoui, H., Tetrault, M.-A., Michaud, J.-B., Pepin, C., Cadorette, J. et Lecomte, R. (2009). The Hardware and Signal Processing Architecture of LabPETTM, a Small Animal APD-Based Digital PET Scanner. *IEEE Transactions on Nuclear Science*, volume 56, numéro 1, p. 3–9.
- Fontaine, R., Michaud, J. B., Belanger, F., Cadorette, J., Leroux, J. D., Pratte, J. F., Robert, S. et Lecomte, R. (2003). Design of a Dual-Modality, Digital Positron Emission Tomography/Computed Tomography (PET/CT) Scanner for Small Animal Imaging. Dans *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Dept. of Electr. Eng. Comput. Eng., Sherbrooke Univ., Que., Canada. volume 1. IEEE, Piscataway, NJ, USA, p. 998–1001.
- Fourmigue, A., Beltrame, G. et Nicolescu, G. (2013). Explicit transient thermal simulation of liquid-cooled 3D ICs. Dans *Design, Automation Test in Europe Conference Exhibition, 2013*. p. 1385–1390.
- Gallivanoni, A., Rech, I. et Ghioni, M. (2010). Progress in Quenching Circuits for Single Photon Avalanche Diodes. *IEEE Transactions on Nuclear Science*, volume 57, numéro 6, p. 3815–3826.
- Garrou, P., Bower, C. A. et Ramm, P. (2008). *Handbook of 3D Integration : Technology and Applications of 3D Integrated Circuits (edited by Philip Garrou, Christopher Bower and Peter Ramm)*, volume 1. Wiley-VCH Verlag GmbH & Co. KGaA, 270 p.
- Garrou, P., Bower, C. A. et Ramm, P. (2008). *Handbook of 3D Integration : Technology and Applications of 3D Integrated Circuits (edited by Philip Garrou, Christopher Bower and Peter Ramm)*, volume 2. Wiley-VCH Verlag GmbH & Co. KGaA, 503 p.

- Ghioni, M., Gulinatti, A., Rech, I., Zappa, F. et Cova, S. (2007). Progress in Silicon Single-Photon Avalanche Diodes. *IEEE Journal of Selected Topics in Quantum Electronics*, volume 13, numéro 4, p. 852–862.
- Goldfarb, M. E. et Pucel, R. A. (1991). Modeling via hole grounds in microstrip. *IEEE Microwave and Guided Wave Letters*, volume 1, numéro 6, p. 135–137.
- Guerrieri, F., Tisa, S. et Zappa, F. (2009). Fast Single-Photon Imager acquires 1024 pixels at 100 kframe/s. Dans *Proceedings of SPIE - The International Society for Optical Engineering*. volume 7249. p. 72490U.
- Imoto, T., Mawatari, K., Wakiyama, K., Kobayashi, T., Yano, M., Shinohara, M., Kinoshita, T. et Ansai, H. (2009). A novel ESD protection device structure for HV-MOS ICs. Dans *2009 IEEE International Reliability Physics Symposium*. p. 663–668.
- Katti, G., Stucchi, M., De Meyer, K. et Dehaene, W. (2010). Electrical Modeling and Characterization of Through Silicon via for Three-Dimensional ICs. *IEEE Transactions on Electron Devices*, volume 57, numéro 1, p. 256–262.
- Kim, Y. J., Joshi, Y. K., Fedorov, A. G., Lee, Y.-J. et Lim, S.-K. (2010). Thermal characterization of interlayer microfluidic cooling of three-dimensional integrated circuits with nonuniform heat flux. *Journal of Heat Transfer*, volume 132, numéro 4, p. 1–9.
- Kim, Y. S., Jun, I. S. et Kim, K. H. (2008). Design and Characterization of CMOS Avalanche Photodiode With Charge Sensitive Preamplifier. *IEEE Transactions on Nuclear Science*, volume 55, numéro 3, p. 1376–1380.
- LaFonteese, D. J., Vashchenko, V. A. et Korablev, K. G. (2009). Breakdown voltage walkout effect in ESD protection devices. Dans *2009 IEEE International Reliability Physics Symposium*. p. 659–662.
- Lee, M.-R. et Kwon, O.-K. (1999). High performance extended drain MOSFETs (ED-MOSFETs) with metal field plate. Dans *The 11th International Symposium on Power Semiconductor Devices and ICs, 1999. ISPSD '99. Proceedings*. p. 249–252.
- Los Alamos National Laboratory (2012). *The Milagro Gamma-Ray Observatory*. <http://www.lanl.gov/milagro/> (page consultée le 11 mai 2012).
- Mather, R. L. (1951). Čerenkov Radiation from Protons and the Measurement of Proton Velocity and Kinetic Energy. *Phys. Rev.*, volume 84, p. 181–190.
- Mita, R. et Palumbo, G. (2008). High-Speed and Compact Quenching Circuit for Single-Photon Avalanche Diodes. *IEEE Transactions on Instrumentation and Measurement*, volume 57, numéro 3, p. 543–547.
- Mitros, J. C., Tsai, C.-Y., Shichijo, H., Kunz, M., Morton, A., Goodpaster, D., Mosher, D. et Efland, T. R. (2001). High-voltage drain extended MOS transistors for 0.18- μm logic CMOS process. *IEEE Transactions on Electron Devices*, volume 48, numéro 8, p. 1751–1755.

- Moore, G. E. (1965). Cramming More Components Onto Integrated Circuits. *Electronics*, volume 38, numéro 8, p. 114–117.
- Niclass, C., Rochas, A., Besse, P. A. et Charbon, E. (2005). Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes. *IEEE Journal of Solid-State Circuits*, volume 40, numéro 9, p. 1847–1854.
- Niclass, C. L. (2008). *Single-Photon Image Sensors in CMOS : Picosecond Resolution for Three-Dimensional Imaging*. Thèse de doctorat, École Polytechnique Fédérale de Lausanne, Lausanne, Vaud, Suisse, 251 p.
- Northend, C. A., Honey, R. et Evans, W. E. (1966). Laser Radar (Lidar) for Meteorological Observations. *Review of Scientific Instruments*, volume 37, numéro 4, p. 393–400.
- Nouizi, F. (2012). *Tomographie optique diffuse et de fluorescence préclinique : instrumentation sans contact, modélisation et reconstruction 3D résolue en temps*. Thèse de doctorat, Université de Strasbourg, Strasbourg, Alsace, France, 162 p.
- Pierre Auger Observatory (2012). *Pierre Auger Observatory studying the universe's highest energy particles*. <http://www.auger.org/index.html> (page consultée le 11 mai 2012).
- Precision Measurement Technologies (2013). *TDC-GPX multifunctional High-end Time-to-Digital Converter*. <http://www.pmt-fl.com/tdc/tdc-gpx.php> (page consultée le 26 juin 2013).
- Ramm, P., Lu, J. J.-Q. et Taklo, M. M. (2012). *Handbook of Wafer Bonding (edited by Peter Ramm, James Jian-Qiang Lu and Maaïke M. V. Taklo)*. Wiley-VCH Verlag GmbH & Co. KGaA, 395 p.
- Rech, I., Marangoni, S., Resnati, D., Ghioni, M. et Cova, S. (2009). Multipixel single-photon avalanche diode array for parallel photon counting applications. *Journal of Modern Optics*, volume 56, numéro 2-3, p. 326–333.
- Rochas, A., Besse, P. A. et Popovic, R. S. (2004). Actively recharged single photon counting avalanche photodiode integrated in an industrial CMOS process. Dans *EUROSENSORS XVI*, Inst. of Microelectron. Microsyst., Swiss Fed. Inst. of Technol., Lausanne, Switzerland. volume A110. Elsevier, Switzerland, p. 124–129.
- Saleh, R., Hussain, S. Z., Rochel, S. et Overhauser, D. (2000). Clock skew verification in the presence of IR-drop in the power distribution network. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, volume 19, numéro 6, p. 635–644.
- Santos, P. M., Casimiro, A. P., Lanca, M. et Simas, M. I. C. (2001). CMOS compatible H_v gate-shifted LDD-NMOS. *IEEE Transactions on Electron Devices*, volume 48, numéro 5, p. 1013–1015.
- Santos, P. M., Casimiro, A. P., Lanca, M. et Simas, M. I. C. (2002). High-voltage solutions in CMOS technology. *Microelectronics Journal*, volume 33, numéro 8, p. 609–617.

- Santos, P. M., Costa, V., Gomes, M. C., Borges, B. et Lança, M. (2007). High-voltage LDMOS transistors fully compatible with a deep-submicron 0.35 μm CMOS process. *Microelectronics Journal*, volume 38, numéro 1, p. 35–40.
- Schwarz, B. (2010). Lidar : Mapping the world in 3D. *Nature Photonics*, volume 4, numéro 7, p. 429–430.
- Shrivastava, M., Baghini, M. S., Gossner, H. et Rao, V. R. (2010). Part I : Mixed-Signal Performance of Various High-Voltage Drain-Extended MOS Devices. *IEEE Transactions on Electron Devices*, volume 57, numéro 2, p. 448–457.
- Sonsky, J. et Heringa, A. (2007). Ultra-flexible, layout-enabled field plates for HV transistor integration in SOI-based CMOS. Dans *19th International Symposium on Power Semiconductor Devices and IC's, 2007. ISPSD '07*. p. 77–80.
- Stapels, C. J., Lawrence, W. G., Augustine, F. L. et Christian, J. F. (2006). Characterization of a CMOS Geiger photodiode pixel. *IEEE Transactions on Electron Devices*, volume 53, numéro 4, p. 631–635.
- Stipcevic, M. (2009). Active quenching circuit for single-photon detection with Geiger mode avalanche photodiodes. *Applied Optics*, volume 48, numéro 9, p. 1705–1714.
- Stoppa, D., Mosconi, D., Pancheri, L. et Gonzo, L. (2009). Single-photon avalanche diode CMOS sensor for time-resolved fluorescence measurements. *IEEE Sensors Journal*, volume 9, numéro 9, p. 1084–1090.
- Streetman, B. G. et Banerjee, S. K. (2005). *Solid State Electronic Devices*, 6^e édition. Pearson Prentice Hall, 581 p.
- Tetrault, M. A., Viscogliosi, N., Riendeau, J., Belanger, F., Michaud, J. B., Semmaoui, H., Berard, P., Lemieux, F., Arpin, L., Bergeron, M., Cadorette, J., Pepin, C. M., Robert, G., Lepage, M. D., Lecomte, R. et Fontaine, R. (2008). System architecture of the LabPET small animal PET scanner. *IEEE Transactions on Nuclear Science*, volume 55, numéro 5, p. 2546–2550.
- The Imagine Team (2011). *Air Cerenkov Detectors*. http://imagine.gsfc.nasa.gov/docs/science/how_12/cerenkov.html (page consultée le 11 mai 2012).
- Tisa, S., Guerrieri, F. et Zappa, F. (2008a). Monolithic array of 32 SPAD pixels for single-photon imaging at high frame-rates. Dans *New Developments in Photodetection, 5th International Conference on New Developments in Photodetection 2008*. <http://ndip.in2p3.fr/ndip08/Presentations/2Monday/A-Midi/126-Guerrieri.pdf> (page consultée le 6 juin 2012).
- Tisa, S., Guerrieri, F. et Zappa, F. (2008b). Variable-load quenching circuit for single-photon avalanche diodes. *Optics Express*, volume 16, numéro 3, p. 2232–2244.
- Tisa, S., Guerrieri, F. et Zappa, F. (2009). SPAD detection head with 32 fully-parallel channels for time-tagging single-photons at 3 μs . Dans *Proceedings of SPIE - The International Society for Optical Engineering*. volume 7222. p. 72221H.

- Tisa, S., Zappa, F., Tosi, A. et Cova, S. (2007). Electronics for single photon avalanche diode arrays. *Sensors & Actuators : A.Physical*, volume 140, numéro 1, p. 113–122.
- Tudor, B., Wang, J. W., Hu, B. P., Liu, W. et Lee, F. (2008). An accurate and versatile ED- And LD-MOS model for high-voltage CMOS IC SPICE simulation. Dans *Technical Proceedings of the 2008 NSTI Nanotechnology Conference and Trade Show, NSTI-Nanotech, Nanotechnology 2008*, Synopsys, Inc., 700 E. Middlefield Road, Mountain View, CA, United States. volume 3. Taylor and Francis Inc, Quebec City, QC, United states, p. 804–807.
- Velodyne Lidar, Inc. (2010). High Definition Lidar : HDL-64E S2. Dans Velodyne Lidar, Inc., *Velodyne - High Definition Lidar*. http://velodynelidar.com/lidar/products/brochure/HDL-64E%20S2%20datasheet_2010_lowres.pdf (page consultée le 3 mai 2012).
- Wu, K. M., Chen, J. F., Su, Y. K., Lee, J. R., Lin, K. W., Shih, J. R. et Hsu, S. L. (2006a). Effects of gate bias on hot-carrier reliability in drain extended metal-oxide-semiconductor transistors. *Applied Physics Letters*, volume 89, numéro 18, p. 183522–1.
- Wu, K. M., Chen, J. F., Su, Y. K., Lee, J. R., Lin, Y. C., Hsu, S. L. et Shih, J. R. (2006b). Anomalous Reduction of Hot-Carrier-Induced On-Resistance Degradation in n-Type DEMOS Transistors. *IEEE Transactions on Device and Materials Reliability*, volume 6, numéro 3, p. 371–376.
- Young, S. (2010). LIDAR in the Driver's Seat. Dans Optics & Photonics Focus, *Optics & Photonics Focus*. http://www.opfocus.org/content/v12/s7/opfocus_v12_s7.pdf (page consultée le 3 mai 2012).
- Zappa, F., Lotito, A., Giudice, A. C., Cova, S. et Ghioni, M. (2003). Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors. *IEEE Journal of Solid-State Circuits*, volume 38, numéro 7, p. 1298–1301.
- Zappa, F., Tisa, S., Tosi, A. et Cova, S. (2007). Principles and features of single-photon avalanche diode arrays. *Sensors & Actuators : A.Physical*, volume 140, numéro 1, p. 103–112.
- Zappa, F., Tosi, A., Mora, A. D. et Tisa, S. (2009). SPICE modeling of single photon avalanche diodes. *Sensors and Actuators : A Physical*, volume 153, numéro 2, p. 197–204.